Detecting, Investigating, and Responding to HIV Transmission Clusters

Division of HIV/AIDS Prevention
Centers for Disease Control and Prevention

Version 1.0
March 2017

Table of Contents

HIV Cluster Guidance Working Group:

CDC Participants: Anne Marie France, Alexa Oster, Cheryl Ocfemia, Nivedha Panneer, Pat Sweeney, Erica Dunbar, Stacey Muckleroy, Phil Peters, Bill Switzer

HRSA Participants: Susan Robilotto, Marlene Matosky

Health Department Participants: Bridget Anderson, New York State Department of Health; Kathleen Brady, Philadelphia Department of Public Health; Mary-Grace Brandt, Michigan Department of Health and Human Services; Heidi Jenkins, Connecticut Department of Public Health; Melissa Miller, Philadelphia Department of Public Health; Analise Monterosso, Texas Department of State Health Services; Jen Reuer, Washington Department of Health; Heather Bronson, Virginia Department of Health

Section 1. What is a transmission cluster?

- A **transmission cluster** is a group of HIV-infected persons (diagnosed and undiagnosed) who have a direct or indirect epidemiological connection related to HIV transmission. Transmission clusters can represent recent and ongoing HIV transmission in a population, where public health interventions could improve care outcomes and prevent new infections. Section 2, 'How can identifying transmission clusters help focus prevention efforts?' describes the importance of identifying transmission cluster for prevention efforts in more detail.
- A transmission cluster represents a subset of an underlying **risk network**. A risk network includes the group of persons among which HIV transmission has occurred and could be ongoing. This network includes persons who are not HIV-infected but may be at risk for infection, as well as the HIV-infected persons in the transmission cluster. Transmission clusters present opportunities for improvement of HIV health outcomes and prevention of new infections in the larger underlying risk network.
- Transmission clusters can be identified through multiple mechanisms:
    - **HIV case surveillance data.** Aberrant increases in HIV diagnoses in a particular geographic area or population could represent recent and ongoing HIV transmission. In areas with low prevalence of HIV (like many rural communities in the United States), transmission clusters could be detected by monitoring variations in reporting of HIV diagnosis data. Improved quality and timeliness of reporting may improve a jurisdiction's ability to detect a recent transmission cluster. It is important to note, however, that an increase in the number of diagnoses may not reflect an increase in transmission. Rather, an increase in reported diagnoses may reflect an increase in HIV testing that has diagnosed people whose infection may be longstanding.
    - **Molecular HIV surveillance data.** Analysis of HIV sequence data reported to HIV surveillance can identify clusters of cases with closely related HIV strains. This method may be particularly useful in identifying transmission clusters that are not detected through other mechanisms. Examples include transmission clusters occurring in an area with a high incidence of HIV infection, those involving multiple jurisdictions, or those in populations for whom partner elicitation data is limited (e.g. due to anonymous partners). Molecular data can also be used to confirm that clusters identified through other mechanisms truly represent a transmission cluster.
    - **HIV partner services and contact investigations.** Partner services staff (referred to as disease intervention specialists (DIS) in many jurisdictions) routinely conduct HIV contact investigations, elicit partner information, and notify partners of their possible exposure or potential risk of HIV infection. Partner services activities can also include prevention counseling, testing for HIV and STDs, and linkage and referral to medical care and other services. As DIS work intensively in local communities, they are positioned to notice unexpected patterns or increases in HIV diagnoses.
    - **Health care providers, health department staff or community members.** HIV transmission clusters are sometimes first detected through astute observations from frontline staff at the health department or clinical providers. Observations like these

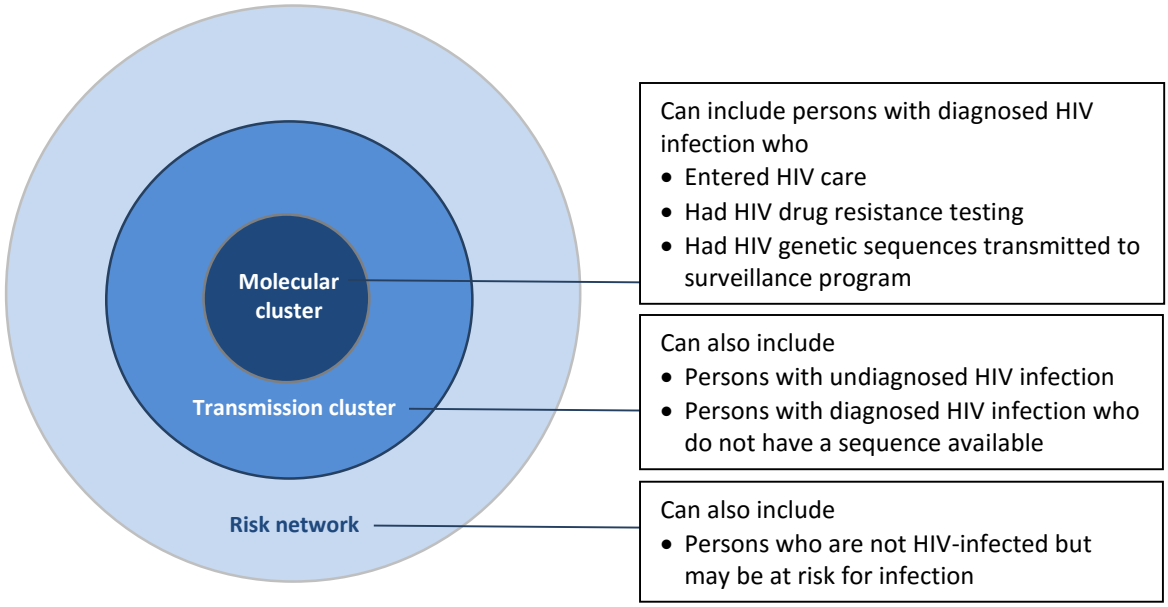
call for further investigation to determine if and how these persons are connected and the extent of other connections they may have in a community.
    - **Changes in patterns of other diseases:** Aberrant increases in diseases that share similar transmission risks such as hepatitis C or STDs could also prompt investigation to identify potential recent and ongoing HIV transmission.

This document focuses on the use of molecular HIV surveillance data to identify transmission clusters for investigation and intervention. Additionally, as our understanding of how to most effectively identify, investigate, and intervene in transmission clusters grows, this guidance will be further developed and refined.

What is a molecular cluster, and how does it relate to a transmission cluster?

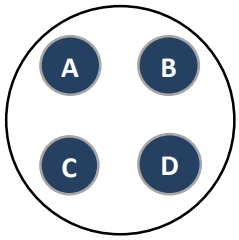
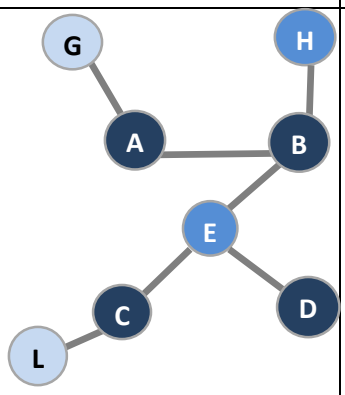
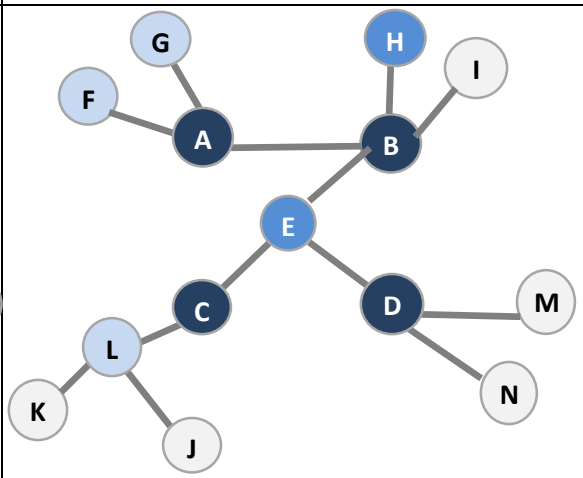
- Identification of **molecular clusters** provides a tool to identify transmission clusters. A **molecular cluster** is a group of persons with diagnosed HIV infection who have genetically similar HIV strains. Because HIV is constantly evolving, persons whose viral strains are genetically similar may be closely related by transmission.
- Molecular clusters are identified through analysis of HIV genetic sequence data that is generated through HIV drug resistance testing. This testing is recommended for all persons with diagnosed HIV infection at entry to HIV care to help HIV care providers select an appropriate treatment regimen. HIV genetic sequence data are routinely reported to the local/state HIV surveillance program and to CDC through the National HIV Surveillance System (NHSS).
- A **molecular cluster** contains only those people for whom HIV genetic sequence data are available and can be analyzed, and contains a subset of what is likely a larger underlying transmission cluster**.**
- As a result, molecular clusters include persons with diagnosed HIV infection who entered HIV care, had HIV drug resistance testing, and their sequences were transmitted to the local/state HIV surveillance program for analysis.
- This represents a subset of the underlying transmission cluster, which can also include:
    - Persons with diagnosed HIV infection who do not have a sequence available for analysis, either because:
        - They did not enter HIV care
        - They entered care, but have not had an HIV drug resistance test
        - They entered care and have had HIV drug resistance test, but the sequence was not transmitted to the health department for analysis, or was of poor quality and could not be analyzed
    - Persons with undiagnosed infection
- In addition to the persons in the transmission cluster, the underlying risk network will include:
    - HIV-negative persons at risk for acquiring HIV

Figure 1a. Molecular cluster and its underlying transmission cluster and risk network.

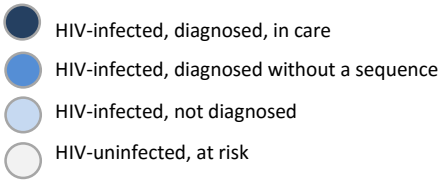


- HIV surveillance sequence data serves as a proxy for an epidemiological relationship but cannot reveal which cases are directly related by transmission or determine the direction of transmission. This limitation is because two persons with genetically similar HIV strains are not necessarily directly linked by transmission: the relationship could be indirect, and there could be unidentified persons involved in transmission relationships.
- Use of HIV sequence data to identify molecular clusters is described in detail in Section 3, 'Identifying growing transmission clusters.'
- Once a molecular cluster is identified, the molecular cluster can be characterized using case surveillance information such as demographic, geographic, risk, clinical and laboratory data; however, the corresponding transmission cluster and risk networks can only be identified through investigation, which includes the assessment of other data sources as described in Section 5, 'Investigating transmission clusters.'

Figure 1b. Hypothetical molecular cluster (a) and corresponding underlying transmission cluster (b) and risk network (c).

a. Hypothetical molecular cluster identified through sequence analysis	b. Hypothetical underlying transmission cluster	c. Hypothetical underlying risk network
		

Legend: Circles represent persons with HIV infection or persons at high risk of HIV infection; lines represent sexual or risk relationships between persons.



Section 2. How can identifying transmission clusters help focus prevention efforts?

- Investigation of transmission clusters can identify risk networks that are concerning for ongoing transmission, early infection, poor health outcomes, or other reasons, such as transmission in a particularly vulnerable or underserved population, or transmission of drug resistance. Networks of concern include:
  - Networks in which HIV transmission occurred rapidly (with multiple new infections occurring within months of one another) and within a recent time window (within ~1-2 years). Recent, rapid transmission suggests the presence of a concerning risk network and could indicate ongoing transmission or a potential outbreak for which public health intervention could interrupt transmission and prevent future infections.
  - Networks characterized by poor outcomes, such as late diagnosis or unsuppressed viral loads; this could suggest poor or limited access to care, and could indicate a network in which persons with HIV infection that has not yet been diagnosed or persons with high viral loads could be contributing to ongoing transmission.
  - Networks representing vulnerable or underserved populations, such as pregnant women, adolescents, rural populations, injection-drug users, foreign-born persons, or other groups defined by local epidemiology and context.

- Networks in which drug-resistant strains of HIV are being transmitted; particularly networks with resistance to first line antiretroviral medications or pre-exposure prophylaxis (PrEP) regimens.
- Networks including persons in stage 0 of HIV infection, which includes acute infection or recent seroconversions (negative HIV test within 180 days of HIV diagnosis), indicate very recent transmission and require immediate investigation.
- Networks not reached by testing efforts, as evidenced by large proportions of cases that were diagnosed through incidental testing, such as screening in plasma centers, emergency departments, or correctional institutions; this could indicate other cases in the network that have not yet been diagnosed and could be contributing to ongoing transmission.

- Investigation of transmission clusters can identify key characteristics of the underlying risk network to guide intervention efforts to improve health outcomes and prevent additional infections.
  - Investigation includes the examination of existing data, including partner services data, or collection of new data, to identify factors associated with transmission. Investigation will be discussed in detail in Section 5, 'Investigating transmission clusters.'
- Intervening in risk networks can improve health outcomes and interrupt transmission through activities that could include:
  - Identifying persons with diagnosed HIV infection in the transmission cluster who are out of care, and ensuring that these cases are linked to or re-engaged in care
  - Identifying persons with undiagnosed infection who are part of the transmission cluster, and linking these persons to care
  - Identifying HIV-negative persons in the risk network who are at risk for acquiring HIV and offering effective prevention interventions, such as PrEP
  - Identifying potential venues, communities, or geographical areas in need of services or broader community interventions.
  - Other transmission cluster and risk network--specific interventions
- By expanding our knowledge of transmission dynamics, transmission cluster data can be a powerful tool to target the interventions we know are effective (engagement in care, HIV testing, PrEP).

Using transmission cluster data to target prevention efforts requires identifying, interpreting, prioritizing, investigating, and intervening in growing transmission clusters. Each of these activities will be described in more detail.

Section 3. Identifying growing transmission clusters

A key strategy to identify growing transmission clusters is through the identification of molecular clusters.

How are molecular clusters identified?

HIV is constantly evolving

- The genetic sequence of HIV accumulates changes over time. Immediately following transmission of HIV between two people, the genetic sequence of the HIV strain in the recipient will be nearly identical to strains found in the transmitting person. As time passes, however, the strains infecting each person will change independently of one another and will look more and more different. In each new person infected, the virus will continue to change independently, so the HIV strains will look less and less similar over the course of a transmission chain. For more detail about the evolution of HIV, including the rate of change, please see Appendix B.
- Analysis of the nucleotide sequence of viral strains can determine how genetically similar the strains are.

- People whose viral strains are genetically similar may be closely related via transmission.

Analysis of sequence data

- Each HIV nucleotide sequences is compared to every other HIV nucleotide sequence to identify pairs of sequences that are extremely similar (i.e., sequences that have a very small genetic distance, or difference). The level of genetic similarity used to identify closely related pairs is referred to as the **genetic distance threshold**.
    - The genetic distance threshold applied can vary based on the goal of the analysis. For example, to identify cases related by recent and rapid transmission, a very close genetic distance threshold should be used—for example, 0.5% (which, for a sequence that is 1000 nucleotides long, corresponds approximately to 5 different nucleotides). A genetic threshold of 0.5% corresponds to approximately a maximum of 2-3 years of viral evolution separating these strains (which may correspond to time since a common transmission event). By contrast, if the goal is to identify all possible cases that could be related to a given case, a larger genetic distance threshold should be used—for example, 1.5%. A 1.5% threshold corresponds to a maximum of 7-8 years of viral evolution separating these strains.
- Pairs of cases with similar sequences are then connected with one another to construct molecular clusters and identify clusters of very closely related cases.
    - Lines are drawn between each pair of sequences that is closely related. This creates clusters that may have as few as two connected sequences, but can contain many more sequences that are connected.
    - Although data on potential transmission linkages between persons (i.e., which pairs of people have genetically linked sequences) are useful in constructing molecular clusters, these data may be subject to misinterpretation such as assumptions that the relationship between two persons is direct (vs. indirect) or that molecular data can establish directionality of transmission, by those not familiar with the limitations of this type of analysis. As a result, CDC recommends minimizing use of these data and instead focusing on cluster-level data (i.e., considering all people in a cluster for intervention rather than focusing on people based on their position in the cluster)
- The time period of data included in the analysis may vary depending upon the goals. Currently, analyses focused on identifying clusters that represent recent HIV transmission include only sequences from cases diagnosed in recent years (e.g., the 3 most recent years).
    - Using a shorter time period, such as 3 years, can identify bursts of recently infected cases that indicate recent and potentially ongoing transmission. Although persons with diagnoses outside of the time window who are out of care could be sources of ongoing transmission, limiting the time window allows the analysis to flag clusters with substantial recent growth. A secondary analysis can then be conducted to identify additional potentially related persons with HIV who should be considered in the investigation.
    - Analysis conducted for other purposes, such as understanding a larger transmission cluster, might include cases diagnosed over a much longer time period.
- For details about the analysis and the software tool HIV-TRACE (HIV TRAnsmission Cluster Engine), including the selection of the genetic distance threshold to define a cluster, a description of the regions of the HIV genome included in the analysis, and other technical details, please see Appendix B.

How are nucleotide sequence data generated and collected?

Generation of nucleotide sequences

- Nucleotide sequences are generated through drug resistance testing.
- The Guidelines for the Use of Antiretroviral Agents in HIV-1-Infected Adults and Adolescents drug recommend resistance testing at entry to HIV care to identify mutations associated with viral

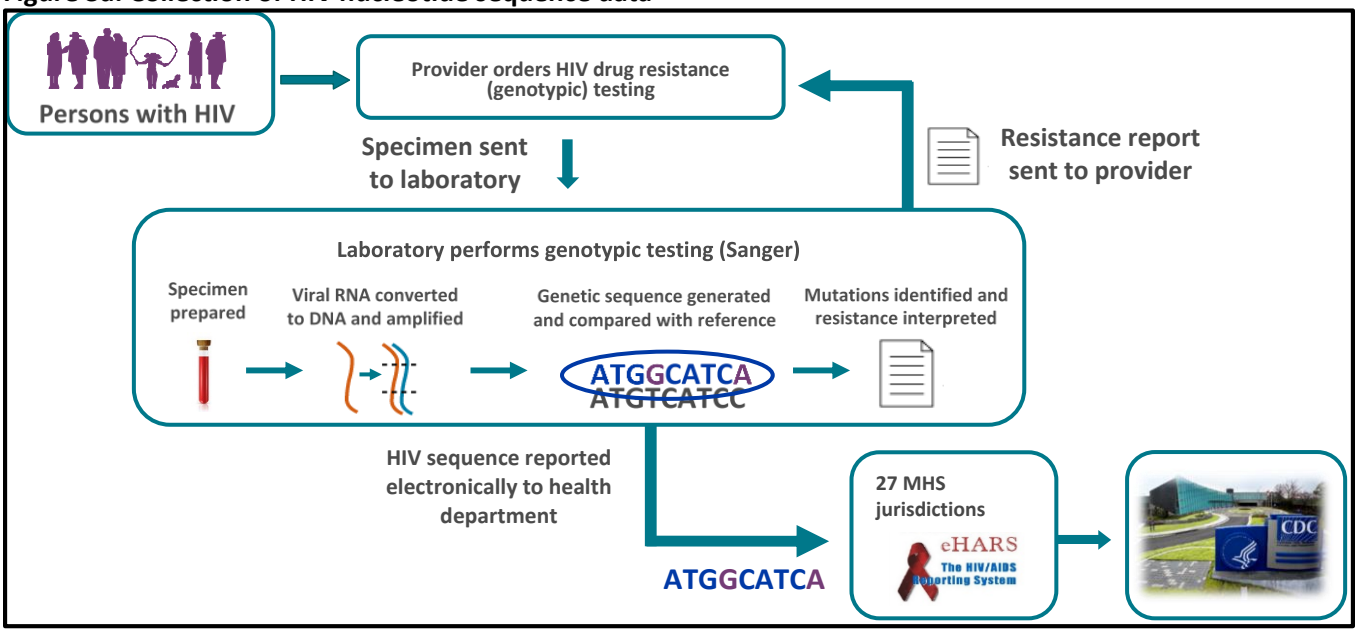
resistance to antiretroviral medications and help the HIV care provider select an appropriate treatment regimen.

- Drug resistance testing is typically ordered by providers at entry to HIV care, but can also be ordered at later time points (for example, if a person is on treatment but does not have a suppressed viral load or when changing ART regimens).
- The final output of drug resistance testing is a report identifying known mutations that confer drug resistance, which is sent to providers. The HIV nucleotide sequence is generated as a part of the testing process, and laboratories can retrieve this information for surveillance reporting purposes. Current testing methods generate sequences using a sequencing method called Sanger sequencing.

Collection of nucleotide sequence data

- Laboratories report HIV nucleotide sequence data to HIV surveillance jurisdictions conducting Molecular HIV Surveillance, an integrated component of the National HIV Surveillance System. During 2013–2017, 27 jurisdictions participate in Molecular HIV Surveillance in the U.S. (see Appendix C for map of jurisdictions).
- Health departments report these data to CDC with all case information collected by HIV surveillance (demographics, transmission category, CD4 results, viral load results) but without identifying information (name, street address) (See Figure 3a)

Figure 3a. Collection of HIV nucleotide sequence data



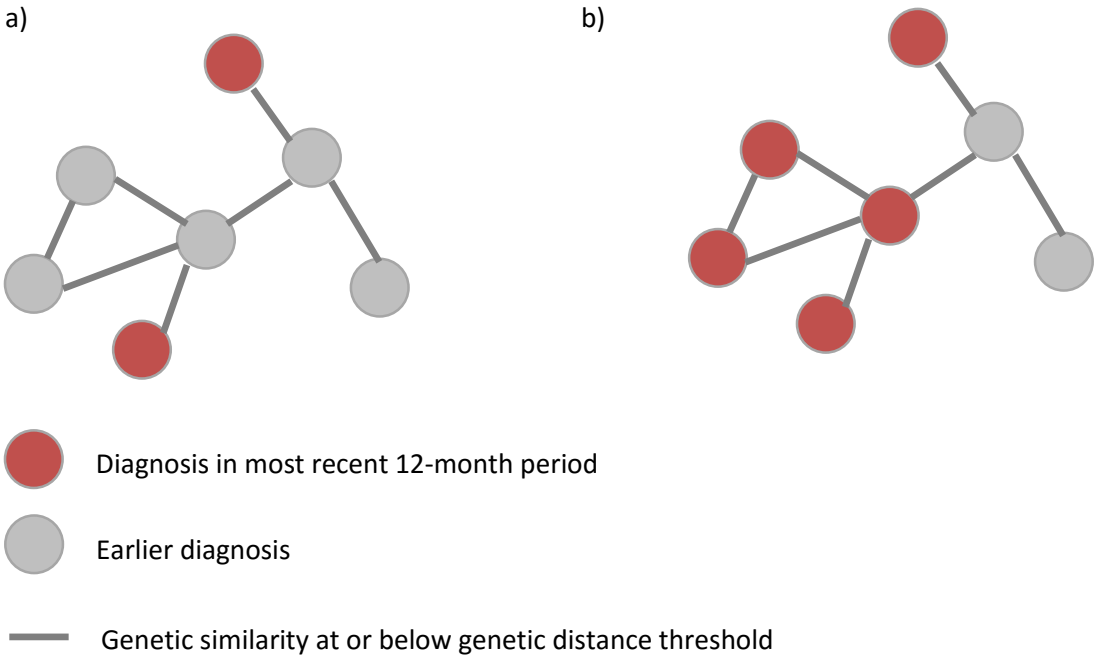
Limitations in nucleotide sequence data

- Although drug resistance testing is recommended for all persons with diagnosed HIV infection, in practice, not all persons receive a drug resistance test at entry to care.
- In some instances, even if a drug resistance test is completed, reporting challenges prevent a health department from receiving a nucleotide sequence for a person.
    - For examples, sequences may not be reported for persons receiving medical care in federal systems (e.g., Veterans Affairs or federal prisons) or those in blinded clinical trials.
    - In some cases, the identifying and locating information provided by the laboratory could be so incomplete that the sequence cannot be linked to a person in the HIV surveillance registry.

How does CDC identify molecular clusters?

- CDC conducts routine analyses to identify molecular clusters that are concerning for recent and rapid transmission of HIV.
- These analyses are conducted using national data that is available each quarter (based on data transmitted by HIV surveillance jurisdictions to CDC using March, June, September, and December datasets).
- Prior to analysis, all HIV sequences in the national dataset are evaluated to determine the quality of the data and remove potential contaminants. Only sequences that include protease or reverse transcriptase regions of the HIV genome and are of sufficient length are included in the analysis.
- CDC analyzes data using a secure local installation of HIV-TRACE, a software tool developed by University of California, San Diego and Temple University.
- With the goal of identifying clusters consistent with recent and rapid HIV transmission, these analyses include only cases diagnosed in the most recent 3-year period, and use a genetic distance threshold of 0.5%.
- Molecular clusters of concern are defined as clusters with at least 5 cases diagnosed within the most recent 12-month period.
- When a molecular cluster of concern is identified, the primary jurisdiction (the jurisdiction with the majority of cases in a cluster) involved is notified and a cluster snapshot describing the cluster is transmitted securely via SAMS. A cluster snapshot companion document, showing the elements included in a cluster snapshot, can be found in Appendix E.
- CDC's prioritization criteria may be modified or expanded in the future, as capacity allows.

Figure 3b. Examples of molecular clusters that would not (a) and would (b) meet CDC's priority criteria.



Can molecular clusters be identified locally?

- Analysis of HIV genetic sequence data by state and local HIV surveillance programs can allow for identification of molecular clusters in closer to real time and for monitoring of clusters that have been previously identified. CDC developed an analytic tool, Secure HIV-TRACE, which allows for state

and local analysis. Secure HIV-TRACE, which will be available in mid-2017, will allow programs to define parameters for molecular cluster identification, such as the genetic distance threshold, to suit analytic goals, while still maintaining the ability to conduct analyses that are in line with CDC approaches.
    - CDC will continue conducting analysis of national data to identify clusters that involve cases from multiple jurisdictions.
- Some jurisdictions have partnered with academic institutions to analyze sequence data to identify molecular clusters. There are many factors that should be considered in such partnerships, including data sharing and security and confidentiality of HIV surveillance information.
- Any use of sequence data for research purposes must be carefully considered and be subject to Institutional Review Board (IRB) approval as appropriate. Any use of identifiable surveillance data as part of any research is contingent on a demonstrated need for the data, IRB approval, and the signing of a confidentiality agreement regarding rules of access and final disposition of the information. Depending on the amount and type of data requested, the use of nonidentifiable data for research is generally permissible but because of the sensitive nature of cluster analyses, IRB approval is recommended and may be required. For more information, see the Data Security and Confidentiality Guidelines for HIV, Viral Hepatitis, Sexually Transmitted Diseases, and Tuberculosis Programs (available at https://www.cdc.gov/nchhstp/programintegration/docs/PCSIDataSecurityGuidelines.pdf).

Section 4. Prioritizing molecular clusters for investigation and intervention

Analysis of HIV sequence data can lead to the identification of large numbers of molecular clusters in a jurisdiction. Molecular clusters can represent HIV transmission clusters that could be concerning for a number of reasons, and not all clusters will be equally concerning from a public health perspective. Prioritizing molecular clusters for investigation and intervention is important to effectively focus public health resources on those clusters on which public health intervention is likely to have the greatest impact.

The process of molecular cluster identification (described in Section 3, 'Identifying growing transmission clusters') is closely related to cluster prioritization, as choices made in the analysis of sequence data can have a large impact on the number and composition of clusters identified. In general, in order to focus on recent transmission, CDC recommends identifying clusters using a smaller genetic distance threshold (0.5%), and limiting analyses to cases diagnosed the most recent 3-year period. Analysis using a larger genetic distance threshold (e.g., 1.5%) can identify clusters where transmission occurred in the more distant past, and where transmission connections between cases are more likely to be indirect. Additionally, analysis conducted using datasets that include cases diagnosed over many years may result in the identification of large, complex clusters comprised of many independent transmission chains, where investigation and intervention could be challenging..

Factors to consider in prioritizing molecular clusters

After molecular clusters are identified, the following factors should be considered in prioritizing clusters for investigation and potential intervention.

- **Potential for ongoing transmission.** Are there factors that raise concern for potential ongoing transmission in the cluster? Such factors include:
    - **Extent of recent cluster growth**
        - How many cases in the cluster have been diagnosed more recently, particularly within the most recent 6–12 month period? Multiple recently diagnosed cases could

indicate ongoing transmission, particularly if evidence suggests that these cases were infected recently.

- **Evidence of early infection**
  - How many cases in the molecular cluster have evidence of early infection ? Such evidence could be acute HIV infections (as determined through the diagnostic testing algorithm), stage 0 infection (as determined through negative testing history), or by a recency result if available. Molecular clusters with evidence of early infection, particularly among recently diagnosed cases, could indicate ongoing transmission.
- **Evidence of cases with unsuppressed viral loads or not in care**
  - How many cases in the molecular cluster do not have evidence of a recent suppressed viral load or evidence that they are in medical care for HIV? Cases not in medical care or not virally suppressed could contribute to ongoing transmission in the cluster. Importantly, cases in the underlying transmission cluster and risk network who are not in the identified molecular cluster could contribute to ongoing transmission, so high levels of viral suppression in the molecular cluster may be falsely reassuring and the potential for such unrecognized cases should be considered.
- **Evidence of ongoing risk behavior that could facilitate HIV transmission**
  - Do cases in the cluster have STD coinfection, or evidence of multiple STD diagnoses? If the cluster involves persons who inject drugs, has hepatitis coinfection been identified in the population at risk? Do partner services interviews note large numbers of sexual partners? These factors could indicate potential for ongoing transmission in the cluster.
- **Extent of the underlying transmission cluster and risk network**
  - Is there evidence that the transmission cluster is larger than molecular cluster based on local data? A review of partner services data from molecular cluster cases could reveal named partners who do not have sequences available for analysis but could be part of the transmission cluster. Additionally, these data could reveal large numbers of unidentified or anonymous partners among molecular cluster cases, suggesting a much larger underlying risk network. This could suggest a high potential for persons with undiagnosed infection contributing to ongoing transmission in the cluster. If partner services interviews were not conducted for cases in the molecular cluster, or if interviews were conducted but no partners were identified, this should raise concern as the extent of the underlying risk network is unknown.
  - By contrast, evidence that all or the majority of cases in a molecular cluster were identified through partner services and that transmission relationships were well understood prior to the availability of sequencing data could be reassuring, as this suggests that the underlying transmission cluster and risk network were captured by frontline public health efforts.

- **Potential for poor health outcomes.** Are there factors that raise concern for the potential for poor health outcomes among cases in the molecular cluster? Such factors could include:
  - **Presence of drug resistance**.
    - Have the same drug resistance mutations been detected among multiple cases in the molecular cluster, suggesting transmission of a drug-resistant strain?
  - **Vulnerable and underserved populations**
    - Is the population impacted by the cluster particularly vulnerable or underserved? Populations of concern could include the very young (e.g., persons aged <20 years),

pregnant women, underserved populations with limited access to health care, or rural or other populations in which testing and treatment facilities are limited. These factors could raise concern for poor health outcomes among cases, as well as the potential for HIV-infected but undiagnosed persons in the transmission/risk network.

- **Late diagnoses**
  - Were many cases in the cluster diagnosed late (e.g., diagnosed with HIV and Stage 3 (AIDS) concurrently, or transitioned to Stage 3 (AIDS) of HIV disease within 6 months of HIV diagnosis)? This could indicate poor access to care among the underlying risk network represented by the cluster.
- **Incidental diagnoses**
  - Were many cases in the cluster diagnosed through incidental testing, such as screening in plasma centers, emergency departments, or correctional institutions? This could indicate other cases in the network that have not yet been diagnosed and could be contributing to ongoing transmission

- **Local epidemiology**
  - Factors relevant to local epidemiology are important to consider in cluster prioritization. For example, does the molecular cluster involve a population of concern based on local epidemiology, or which is a local priority for other reasons?
  - Is there evidence of non-traditional demographic or risk profiles among cases in a certain geographic area? A marked increase in a particular transmission risk category or demographic group may indicate transmission into a novel population. For instance, an increase in cases occurring in young women in an area in which diagnoses typically occur among men who have sex with men likely warrants further exploration.
- **Opportunities to intervene**
  - If transmission in a cluster occurred in the distant past, opportunities to intervene might be limited. For this reason, timely identification of transmission clusters is critical.
    - The addition of new cases to a molecular cluster could indicate new transmission and potential opportunities to intervene; therefore, continued monitoring of molecular clusters is important.

Prioritization is a dynamic process, and factors used for prioritization could change as new information becomes available or new cases are added to a cluster.

Data for molecular cluster prioritization

Cluster prioritization can be initiated using data that is readily available at the health department. Such data can include local HIV surveillance data, STD surveillance data, partner services data, and other locally available data sources. The more data considered, the more refined a prioritization approach will be.

CDC's prioritization approach

As described in Section 3, 'Identifying growing transmission clusters', CDC's current molecular cluster identification approach focuses on identifying a small subset of molecular clusters using quarterly national data sets that are most concerning for recent, rapid transmission and growth that could represent an outbreak. The clusters of concern are currently selected based on clustering at a low genetic threshold (0.5%) that is consistent with recent transmission (choosing a low threshold identifies pairs of sequences that are very closely related, meaning that there has not been time for the virus to evolve much between transmissions).

Among molecular clusters meeting this threshold, CDC prioritizes those clusters that have at least 5 new cases in the cluster diagnosed in the most recent 12 month period, indicating rapid growth. CDC notifies jurisdictions of all clusters meeting these priority criteria. CDC's prioritization criteria may be modified or expanded in the future, as capacity allows.

Developing program capacity to identify and prioritize molecular clusters for investigation

Routinely identifying and prioritizing molecular clusters for investigation will require clear processes and staff. Developing such a program is described in detail in Section 10, 'Creating program capacity for cluster detection and response.'

Resources for investigation and intervention

Resources for cluster investigation and intervention will vary across jurisdictions. If concerning molecular clusters are identified in jurisdictions where resources for investigation and intervention are limited, other resources should be considered. For example:

- CDC may be available to provide technical assistance for a cluster investigation. Such assistance could either be remote or onsite. To discuss potential technical assistance, please contact the team lead of the HIV Incidence and Case Surveillance Branch's Incidence and Molecular Epidemiology Team (Alexa Oster, aoster@cdc.gov) and your CDC-assigned epidemiologist.
- Other jurisdictions might have expertise or investigation resources (such as chart review or interview tools) that could be shared; CDC can facilitate the exchange of these materials.
- There may be opportunities to use partnerships with community-based organizations for some aspects of an intervention, such as field-based testing. Please see Section 10, 'Creating program capacity for cluster detection and response,' for more discussion on this topic.
- Involving care partners early in an investigation might help identify opportunities for support.

Section 5. Investigating transmission clusters

What is a cluster investigation?

A cluster investigation is a systematic process to:

1) Identify the underlying transmission cluster and risk network (e.g., undiagnosed cases, diagnosed cases without a sequence, persons at risk for HIV).
2) Identify factors possibly associated with transmission.
3) Understand possible connections between cases in a cluster.
4) Assess potential risk for ongoing transmission.
5) Determine what potential interventions might be effective.

The first steps of a cluster investigation can be conducted with readily available data, described in more detail below. At each step in the process, new information should be reviewed and assessed, and the cluster can be re-prioritized to assess further investigation as needed. In some cases, a review of readily available information on the cluster might be sufficient, if it establishes that the level of concern for ongoing transmission or unrecognized cases is low. In other cases, collection of additional information, including potential review of medical records and interviews, might be needed.

Although traditionally it is expected that investigation precedes intervention, in reality, these two activities may overlap temporally. For example, certain interventions may be so critical and universally called for that they should be initiated while investigation is ongoing. This investigation can then inform additional intervention needs. Figure 5 provides a roadmap to investigating and intervening.

Figure 5. Roadmap to investigating and intervening in transmission clusters

Step 1	**Ascertain known transmission cluster and risk network**
	• Use partner services data to identify all partners of molecular cluster members (ideally, also identify partners of the partners). Creating a network diagram of the transmission cluster can be useful.
	  • For HIV-positive persons in the transmission cluster, use surveillance and other data sources to determine updated viral suppression information and early infection status
	  • For HIV-negative persons in the risk network, determine time since most recent HIV test and whether person was referred for PrEP
	  • For all persons identified, determine whether/when partner services/testing was conducted and the disposition
	
	More detail available in Section 5, 'Investigating transmission clusters' and Appendix F, 'Suggested variables to be captured during a cluster investigation.'

Step 2	**Initiate critical initial interventions** • Initiate engagement in care efforts for all HIV-positive persons who do not have evidence of recent (past 12 month) viral suppression, particularly those with early infection • Consider also re-initiating partner services interview to elicit partners during time since previous partner services interview while not virally suppressed • Initiate partner services for persons for whom an HIV partner services interview was not previously conducted More detail available in Section 6, 'Intervening in transmission clusters and risk networks.'
Step 3	**Determine additional data needs** • Identify available sources of data and variables to be captured to identify factors possibly associated with transmission and determine risk for ongoing transmission • Collect and analyze data More detail available in Section 5, 'Investigating transmission clusters' and Appendix F, 'Suggested variables to be captured during a cluster investigation.'
Step 4	**Implement additional interventions, as appropriate** • Conduct social network approaches to testing, including using incentives and evaluating and referring persons who test negative for PrEP • Consider additional interventions depending on the characteristics of the cluster, which could include but are not limited to targeted outreach at venues, communication campaigns through media or apps, communicating with providers, health alerts, and scale-up of PrEP or testing services More detail available in Section 6, 'Intervening in transmission clusters and risk networks.'

Components of a cluster investigation

1) Create a case definition

The first step of a cluster investigation is to create a case definition for the cluster. Since a molecular cluster identified through analysis will include only those cases that are diagnosed, linked to care, and have had a genetic resistance test with the result submitted to the health department, a case definition is the first step toward defining the likely true extent of a transmission cluster.

A case definition should capture not only those cases that were identified through molecular sequence analysis, but also cases without sequences available that could be part of the underlying transmission cluster. A case definition is typically tiered, and includes categories for cases that are confirmed cluster cases (e.g., cases identified in the molecular cluster), as well as cases that do not have sequences but could be part of the cluster (e.g., cases named as partners to another cluster case, or social network contacts of a cluster case)

Example Case Definition:

- Confirmed cluster case: HIV-positive persons who are linked to the cluster through their HIV sequence
- Probable cluster case: HIV-positive persons who are named partners of confirmed cases
- Possible cluster case: HIV-positive persons who are in the social network of confirmed cases OR HIV+ persons who are named partners of probable cases

Cases outside of your jurisdiction. Transmission clusters and risk networks often cross jurisdictional borders, and investigations should not stop at these borders. Additionally, some persons may reside in one jurisdiction and seek care in a different jurisdiction. A case definition and subsequent investigation should include all persons linked to a cluster, regardless of jurisdiction. For details on coordinating investigations across jurisdictions, see Section 9, 'Approaching multijurisdictional clusters.'

2) Identify available sources of data

A cluster investigation can begin with readily available data. These data can include, but are not limited to:

a. Local HIV surveillance data
b. Partner services data, including from other STD diagnoses
c. Local STD and viral hepatitis data
d. Ryan White HIV/AIDS Program data, including AIDS Drug Assistance Program (ADAP) data
e. Data from other jurisdictions, including HIV surveillance and partner services data
f. Correctional databases
g. Social network sites
h. Discussions with disease intervention specialists (DIS) that interviewed cluster cases

3) Prepare to organize and visualize information generated

A cluster investigation will generate a large amount of information, and developing a database to input, organize, and maintain this information will be critical to effectively assessing the data. Simple excel spreadsheets may be sufficient for most investigations, though a more complex relational database could also be useful, particularly when managing data on connections between persons. At a minimum, a database should include one row of data for each case that meets the cluster case definition, and columns for each of the variables. Additional tables can be used for partners of persons meeting the cluster case definition. Suggested variables and tables to consider are described in Appendix F, 'Suggested variables to be captured during a cluster investigation.'

a. **Consider what variables you will capture for each case.** Variables can always be added to a dataset, but consider starting with a template based on key variables from readily available datasets.
    i. HIV surveillance data can form the backbone of the dataset. Variables to consider including from HIV surveillance include state number, age, sex, race/ethnicity, address, transmission category, recency information (e.g., recency testing), date of last negative HIV test, facility of diagnosis, most recent viral load, drug resistance information, provider that ordered the most recent laboratory tests, and any other key variables that are available and relevant.
    ii. Create a list of variables that you can pull from each additional data source.
b. **Consider how you will analyze and visualize data.**
    i. Microsoft Excel could be adequate for analysis for smaller clusters. Alternatively, data can be imported from Excel into SAS for analysis.
    ii. Visualization tools will be most useful for illustrating linkages between cases, or connections to common venues. Multiple tools could be used for this purpose, including Microsoft Visio and NodeXL (a free add-in for Microsoft Excel). For smaller clusters, manually drawn networks in Microsoft PowerPoint or Word could be sufficient.

4) Systematically gather and review readily available data to address key investigation questions

The next step is to systematically review the data, starting with reviewing partner services data to identify cases that meet the case definition.

a. **Review partner services data to identify cases meeting the case definition and assess the size of the underlying risk network** (e.g., undiagnosed cases, diagnosed cases without a sequence, persons at risk of HIV)

    i. For each case in the identified molecular cluster, review partner services data to identify named partners and named social network contacts who meet the case definition for cluster inclusion. If the case definition includes a category that captures partners of partners, continue this process by identifying named partners of the partners of confirmed cluster cases.
        1. In some cases, screen names or handles on internet sites or apps might be available (perhaps in interview notes). If so, consider including this information as it could provide a mechanism to make connections between cases and identify additional partners. Additionally, some programs are able to use this information to reach out to contacts.
        2. Consider mapping the network to illustrate connections with tools such as Visio or NodeXL.
    ii. Were cluster cases named as partners to other cluster cases? If so, this could be reassuring that the network has been captured well through partner services. If not, however, this suggests that the network has not been captured through partner services investigation, and raises concern that the underlying transmission/risk network could include additional HIV+ infected partners that have not been diagnosed, or HIV- partners at high-risk of HIV infection.
        1. Consider that a retrospective review of cases might offer opportunities to recognize connections between cases that weren't identified during DIS interviews. This could include connections that span jurisdictions, where the state perspective might help to make connections. For details on coordinating investigations across jurisdictions, see Section 9, 'Approaching multijurisdictional clusters.'

If cluster cases have not named each other as partners, are there social network connections between cases? Resources to identify social connections between cases that weren't identified through partner services could include social networking sites such as Facebook, people searching tools such as Lexis-Nexis/Accurint, and apps and internet hookup sites, assuming screen names are known.

iii. Among partners identified that tested negative, were they evaluated and referred for PrEP?

iv. Have persons in the cluster not been interviewed, refused interview, refused to name partners, or cited anonymous partners? This suggests an unknown underlying transmission cluster and risk network that could potentially be much larger than identified through available data, which might elevate the level of concern.

v. Are there geographic connections between new infections? For example, do new diagnoses have a link to recent release from a particular correctional facility or residential care facility?

vi. Are there other connections between new infections? For example, are there multiple pregnant women, with or without men included in the cluster?

b. Identify factors possibly associated with transmission

i. From partner services data, have common venues been named? Are apps/internet sites frequently mentioned? Are sex parties or other gatherings noted? Do persons report high risk behaviors such as anonymous sex or injection or other drug use? Reviewing notes from partner services interviews will likely be needed to answer these questions.

1. Is there geographic similarity in venues/hook-up sites? Consider that there could be geographic connections with these sites even if there is not geographic similarity with residential addresses.

ii. From STD data, were cluster cases previously or currently infected with other STDs? STD diagnoses prior to the HIV diagnosis indicate provider encounters that could have been missed opportunities for provision of or referral for services such as PrEP.

iii. From surveillance, care, or ADAP data, is there evidence of delays in diagnosis (e.g., AIDS diagnosed at or within 6 months of diagnosis) or limited access to medical care? Are persons currently in medical care? Are there persons who are in medical care but have long periods of unsuppressed viral loads or currently unsuppressed viral loads?

iv. From surveillance or other data sources on location, analyze data at finer geographic levels and assess prevention and care services in this location. Determine locations of diagnosis and care to determine whether collaborations with certain providers would be useful.

c. Assess potential risk for ongoing transmission

i. Was the timing of infection likely recent or distant? Recent infection suggests that transmission were recent and may be ongoing and indicates a potential to intervene to prevent new infections. To address this question, consider:

1. Stage of infection at diagnosis – determine whether cases had acute/stage 0 infection, or late diagnosis

a. Stage of infection might be adequately captured by surveillance data, particularly if reporting of the full diagnostic algorithm is good and if data on last negative HIV test is well captured. Alternatively, additional data sources could provide information

to establish likely timing of infection. This could include alternative sources of data on last negative HIV test (e.g., data from STD programs or HIV testing programs), or medical chart review. Data indicative of diagnosis during acute infection or symptoms consistent with acute infection could also be identified from medical chart review.

2. Dates of diagnosis. Even if many cases were diagnosed with acute infection, if the cases were diagnosed in the distant past, this is not necessarily evidence of current ongoing transmission. However, delays in reporting of both case and sequence data should be considered in this assessment.

ii. Are risk behaviors that increase the likelihood of ongoing HIV transmission common among cluster cases? To address this question, consider:
    1. STD diagnoses suggesting ongoing risk behavior. From a review of STD data, have cases been diagnosed with STDs since their HIV diagnosis?
    2. Evidence of sexual risk behavior captured in HIV and other STD partner services interview records and notes. Risk behavior to consider could include:
        a. Anonymous partners
        b. Multiple partners
        c. No or little condom use
        d. Use of drugs during sex
        e. Public sex environments, such as bathhouses and bookstores
        f. Sex parties
        g. Transactional sex
        h. Human trafficking or other forms of victimization
    3. Evidence of injection drug use. This could be captured in partner services interview data or medical record data. Alternatively, linkages with viral hepatitis data could identify hepatitis C virus coinfection suggesting injection drug use.

iii. Are there persons in the cluster that are currently out of medical care or are in medical care but have unsuppressed viral loads?
    1. Persons out of medical care present an ongoing risk, and present opportunities for linkage to medical care. Persons who have had lapses in adherence, even if currently in medical care, might need to be re-interviewed by partner services if there was a period of time when they were not virally suppressed and for which partners were not captured in previous partner services interviews.
    2. Sources of data on care and viral load data include:
        a. Surveillance data (reported CD4 and viral load tests)
        b. Ryan White HIV/AIDS Program data
        c. ADAP
        d. Medical chart review

iv. Does available evidence suggest persons in the underlying risk network are actively seeking testing or are effectively reached through routine testing programs, is there evidence that persons in the underlying risk network aren't being captured rapidly through testing? Evidence that persons aren't being captured rapidly through testing raises concern that multiple HIV-positive persons in the cluster might remain undiagnosed. To address this question, consider:

1. Facility of diagnosis and reason for testing:
    a. Were persons in the cluster identified through partner services or routine, self-initiated testing?
    b. Were persons identified through passive screening that suggests that they aren't seeking routine testing or being identified through partner services? This could include:
        i. Plasma centers
        ii. Screening in emergency departments or other medical settings
        iii. Jail intake screening
        iv. Testing for insurance purposes
        v. Military intake or other routine screening
2. Frequency of testing
    a. Is there evidence that persons in this transmission/risk network test frequently? Consider documented or self-reported data on previous negative HIV tests. Data on negative HIV tests might not be routinely available, however, potential sources of this information could include:
        i. STD data on HIV testing through DIS
        ii. Medical chart review
        iii. Testing/outreach program data (though this might be less likely to be available)
3. Have cases been diagnosed through partner services or self-initiated routine testing? Or have they been diagnosed through required screening, e.g. jail intake or plasma donation centers?

5) **Synthesize data**

Once data has been systematically gathered for all persons meeting the case definition (or the portion for which data is available), the next step is to systematically review this data to address the questions outlined in the above sections, identify gaps that remain that would warrant the collection of additional data, and update the priority level based on the information available to determine if additional investigation or intervention is needed.

a. Systematically review and analyze data
    i. Review the data for all persons meeting the case definition to characterize what is known about the transmission cluster, identify any commonalities between cases, and address the key questions outlined in the above sections and other key questions determined to be important to the investigation. This review could include both quantitative and qualitative review of the data, such as simple descriptive statistics for systematically collected variables, and a qualitative assessment of themes identified in partner services notes.
        1. Simple quantitative analysis could be conducted in MS Excel, or data could be imported to SAS or other analysis tools for analysis.
    ii. Construct diagrams of identified connections between cluster members to visualize what is known about potential transmission relationships between persons in the cluster.
    iii. Consider constructing brief narratives for each person in the cluster. These narratives could contain key demographic, social, and risk information related to the person, as well as presenting the clinical story from any initial symptoms or presentation to testing, to the diagnostic course and care/treatment history. Such narratives can help identify recurring issues and themes across the transmission cluster and may help identify potential intervention points.

Additionally, such narratives can be compelling to tell the 'story' of the cluster, and may help communicate key issues to leadership and stakeholders. While narratives could be initiated from review of available surveillance and partner services data in many cases, these narratives could be further developed by adding information from medical chart reviews and interviews or re-interviews, if conducted.

Example narrative

- 32 year old uninsured gay Hispanic male who lives with his parents; college graduate, was unemployed at time of diagnosis.
- Reports sex only with men, and notes anonymous sex, both insertive and receptive anal sex, and over 200 lifetime partners. Meets partners via GRINDR and at a local festival.
- Diagnosed in 5/2015 by an STD field DIS indicating that he was named as a partner by someone else; previous self-reported negative was in 10/2014. No acute symptoms reported. No STD diagnoses.
- Was interviewed by partner services for HIV diagnosis, claimed 200 anonymous partners in the 12 months prior to diagnosis; no named partners.
- Was linked to care at University Clinic on 6/2015 and started treatment in July 2015. First viral load demonstrating suppression was 12/2015, and no unsuppressed VLs after that time. Most recent visit was 6/2016, with a suppressed viral load in 5/2016.

iv. Consider constructing a narrative that describes the cluster. This can help to summarize key findings, which may be particularly useful in reviewing cluster investigations or if a cluster begins to grow again at a later time.

b. Assess whether key questions can be adequately addressed with the readily available data that was collected, or if there is a need to gather additional data.

i. Is there sufficient information to assess and update the level of concern for ongoing transmission and potential for undiagnosed cases? If so, this information can be used to determine whether additional investigation and potential intervention is needed. If there is not sufficient information to address this question, more information should be gathered.

ii. If key questions remain about how or why persons were seeking testing, frequency of testing, likely timing of exposure, or treatment and viral load history, review of medical charts could be warranted.

1. Medical chart review can also be an important source of information on symptoms of acute HIV infection and missed opportunities for diagnosis of HIV, which could be useful to understand community and systems issues contributing to ongoing transmission.

iii. If key questions remain about partner history (for example, if persons were not interviewed by partner services, or if there were periods during which persons were not virally suppressed for which partners were not elicited), risk behaviors, and meeting venues, interview or re-interview, where possible, should be considered. Enhanced interviews can include more specific or in-depth questions that can be useful in understanding a cluster. Interviews with front-line DIS staff, if not already completed, could also be helpful to address these questions.

iv. In addition to medical chart review and interview or re-interview, additional sources of information to consider include:

1. Discussions with person's HIV care provider
2. Discussions with Ryan White medical case manager

6) **Other additional data to consider**

a. In some cases, collection of specimens for persons who do not have nucleotide sequences available could be useful for an investigation, in order to rule probable and possible cluster cases in or out of the cluster. In some situations, genetic resistance testing might have already been conducted for cases for which sequence data has not been received by the health department; in these cases, working with laboratories to ensure this data is reported will likely be more efficient than collecting specimens for sequencing.

7) Begin identifying intervention opportunities

Even if additional data collection will occur, there could be opportunities to begin interventions to interrupt transmission once readily available data has been collected and reviewed. Such intervention opportunities could include both individual-based and general interventions, as described in Section 7, 'Intervening in transmission clusters.'

8) Continue to synthesize information and refine intervention strategies as new information becomes available

As new data is gathered and synthesized, understanding of factors driving transmission and determination of effective strategies for intervention can be refined. Continue assessing new information as it becomes available, including adding new cases to the case definition as they are reported and assessing information for these cases.

Requesting CDC assistance

- CDC may be available to provide remote or on-site technical assistance with cluster investigation and intervention
- If you are interested in exploring this option, please contact Alexa Oster (aoster@cdc.gov) and your CDC epidemiologist

Section 6. Intervening in transmission clusters and risk networks

Implementing prevention interventions

Interventions seek to interrupt transmission, therefore preventing future infections, and to address factors that facilitated transmission, therefore reducing the likelihood of future clusters in the same population or risk network.

Prevention interventions should always be determined based on the characteristics of people in the cluster and the factors facilitating transmission, as determined through cluster investigation. Still, some interventions, such as efforts to re-engage persons who are not in HIV care, should not be delayed until the completion of a cluster investigation (see Figure 5, 'Roadmap to investigating and intervening in transmission clusters,' located in Section 5, 'Investigating transmission clusters'). Interventions must always be tailored to the jurisdiction and specific characteristics of the population impacted; the guidance that follows is therefore intended as a set of general recommendations, to be tailored appropriately. Additionally, as our understanding of how to most effectively intervene in transmission clusters and risk networks grows, this guidance will be further developed and refined.

Planning ahead for interventions

Planning for interventions in transmission clusters should ideally begin before any cluster is identified. Identifying available resources in advance is critical to ensuring quick intervention when clusters are

detected, making preparations in advance for cluster intervention and resources that would be needed. For guidance on preparations that should be considered in advance of any needed response to a growing transmission cluster, see Section 10, 'Creating program capacity for cluster detection and response.'

Types of interventions

Interventions for growing transmission clusters can take two main forms: individual-level interventions, and structural or population-level interventions.

- **Individual-level interventions:** The primary type of intervention for a risk network is individual-level interventions. The core component of individual-level interventions is to identify persons involved in the risk network and make provisions to reduce transmission to or from these persons, including ensuring that HIV infections are diagnosed, that persons with HIV are engaged in medical care and virally suppressed, and that HIV-uninfected persons are evaluated and referred for PrEP and other prevention services. It is critical that individual-based interventions address the entire risk network, not just cases identified as clustered through sequencing.
    - **Identifying persons involved in the transmission cluster:** Rigorous cluster investigation is the first step to identify all persons in a transmission cluster, but it will often not be adequate to identify the full transmission cluster and risk network. Therefore, a major component of intervention will typically be **case finding** and **partner/risk-network identification.** Strategies to consider for identification of cases and the partner/risk network are described in **Table 1.**
        - As new persons with HIV in the transmission cluster are identified, include these persons in cluster investigation and assess factors that prevented earlier recognition, as these factors could help in the identification of other unrecognized cases or at-risk partners

Table 1. Strategies to identify previously unrecognized partners for testing and provision of prevention interventions

Cluster characteristic	Strategy
All clusters	Initiate partner services for cases for which interviews were not already conducted Re-initiate partner services for cases with periods for which they were out of care or not virally suppressed Conduct social network approaches to testing, including using incentives and evaluating and referring persons who test negative for PrEP. For more information about social network testing strategies, see Appendix D, 'Additional resources.'
Clusters associated with a particular physical venue	Conduct targeted outreach at identified venues for testing and PrEP evaluation and referral; consider partnering or coordinating with community-based organizations that might already be actively involved in working with these venues
Clusters associated with internet sites and apps, particularly where partners are often anonymous	Have partner services explore reaching partners through apps and work with clients during partner services interviews to identify screen names or handles for partners whose name

	they do not know; also consider strategies to determine whether these persons came in for testing. Consider general messaging campaigns through apps or internet sites to encourage testing and prevention strategies, while keeping in mind that these sites are large, and more nuanced targeting (e.g., based on geography) might be needed to more effectively reach the network involved. Consider partnering with community-based organizations that have experience working through apps and internet sites

- **Linkage to medical care for persons in the transmission cluster who are out of medical care:** An outcome of a cluster investigation should be the identification of persons with diagnosed HIV in the transmission cluster who are out of medical care, or in medical care but not virally suppressed. The identification of persons with diagnosed HIV who are out of medical care should be conducted for the entire transmission cluster, and not merely for those cases who were identified as part of the cluster based on sequence data.
  - Consider using existing data-to-care programs to provide linkage for persons identified in a growing transmission cluster
- **Evaluation and referral for PrEP and other prevention services for persons in the transmission/risk network who are HIV uninfected but at risk**
  - Persons identified in the risk network of a growing cluster who are tested and found to be HIV negative are likely at high risk of HIV infection, and could benefit from evaluation for PrEP eligibility and, if eligible, referral for PrEP and other prevention services
  - In some situations, existing PrEP services might be insufficient to provide PrEP services for persons identified in a risk network. In these situations, consider what resources are available to increase availability of PrEP services.
- **Additional services that could be important for persons in the transmission cluster and risk network**
  - Testing for sexually transmitted infections and hepatitis C
  - Vaccination for hepatitis A, hepatitis B, and/or human papillomavirus
  - Referrals to syringe services programs, mental health services, or substance abuse treatment programs, as appropriate
  - Referral for ADAP, Medicaid, or other health insurance and medication assistance programs
  - Referral for other healthcare and prevention services, including sexual health and family planning, as appropriate
  - Referral for social services, including housing, food assistance, and transportation, as appropriate

- **Population-level interventions:** In addition to individual-level interventions, a cluster investigation may reveal structural or process factors that contributed to transmission, which could be addressed through population-level interventions. While population-level interventions might overlap with general prevention interventions already being undertaken in a jurisdiction, information from cluster investigations can provide more focused understanding of the gaps of current prevention programs, and specific populations that are not being effectively reached by

these activities. Additionally, growing transmission clusters could provide justification for the expansion of population-based interventions, or implementation of interventions which had been considered but not yet implemented.

- Population-based interventions will vary based on the specific factors and circumstances identified through cluster investigation. **Table 2** provides examples of population based interventions and scenarios in which they might be warranted. Many other population-based interventions might be appropriate depending on the circumstances of a particular jurisdiction and transmission cluster.

Table 2. Examples of population-based transmission cluster interventions, and situations in which they might be warranted.

Situation	**Population-based intervention**
Risk network includes many persons at risk of HIV infection, or who had provider encounters identifying high-risk behaviors prior to HIV infection, but were not on PrEP	Expand PrEP resources available to affected population Identify opportunities for PrEP evaluation and referral that reach affected network, such as STD clinics Conduct media campaigns to increase PrEP awareness Conduct targeted outreach to providers in areas where clusters have occurred to educate about PrEP
Cases in transmission cluster presented for care during likely acute HIV infection, but were not diagnosed	Provide provider education on the HIV diagnostic testing algorithm Disseminate health alerts to providers alerting them to rapid HIV transmission in the area and the need for HIV testing, increased recognition of acute HIV symptoms, and completion of HIV diagnostic testing algorithm
Cluster in a remote location with few providers trained in HIV testing and treatment	Provide provider education on HIV testing and treatment
Area with insufficient HIV testing resources	Consider short-term and long-term strategies to increase access to HIV testing Consider provision of home testing kits
Infrequent HIV testing in affected risk network	Increase testing resources Incentivize testing Conduct media campaign or other messaging to affected community about risk of HIV transmission and importance of HIV testing
Partner services not initiated for large portion of cases in the HIV transmission cluster	Consider protocol or policy changes to increase partner services and/or prioritize partner services

	for populations involved in rapidly growing transmission clusters
High percentage of cases not located or refused interview, or low partner indexes in interviews or re-interviews, leading to few or no members of cluster being identified or linked	Consider additional training for DIS, based on identified factors leading to unsuccessful attempts to interview, to improve outcomes Retrain DIS to improve partner elicitation skill; consider cultural competency issues; investigate other issues leading to poor elucidation of sociosexual networks
Apps and internet hookup sites important in risk networks, with insufficient information provided to DIS staff in partner services interviews to identify partners	Address barriers to DIS staff accessing apps and internet sites in order to use these sites for locating partners and messaging to persons in high-risk networks
Insufficient public health response capacity, including DIS staff in an area	Consider approaches to bring additional public health and DIS staff to the affected area (see Section 10, Creating program capacity for cluster detection and response)
Cluster associated with injection drug use	Ensure/expand appropriate provision of syringe exchange and substance use treatment

Partnerships for intervention

- Partnerships will likely be a critical component of intervention strategies, as they present an opportunity to leverage expertise and resources of other groups to effectively direct intervention. A strong partnership between HIV surveillance and prevention programs should be at the core of all cluster prioritization, investigation, and intervention work. In addition, organizations to consider partnerships with could include:
    - **Community-based organizations (CBOs)**
        - CBOs might have extensive experience working with populations affected by a growing transmission cluster, and could have effective strategies to reach these populations. Additionally, CBOs might have access to or experience with strategies for intervention, such as use of apps or internet sites to reach partners, or use of social network testing strategies.
    - **HIV care providers**
        - HIV care providers could be an important component of intervention strategies, helping to access persons in the cluster who entered medical care but have not remained engaged in medical care, or who are engaged in medical care but not virally suppressed
    - **STD clinics**
        - STD programs will likely be involved early in the cluster investigation process, given the importance of STD program information to understanding transmission clusters and risk networks. STD clinics could be an important component of intervention strategies, as a key point to access at-risk populations
    - **Jails and other correctional institutions**

        - In some situations, jails and other correctional institutions might be an important partner for both investigation and intervention. Situations where correctional institutions might be particularly important partners include transmission clusters and risk networks involving injection drug use, other substance abuse, or sex work.
    - **Other organizations, as appropriate.**
- When considering partnerships, consider that it will likely be easier to create the foundation for strong partnerships before any transmission cluster is identified. For more information on planning ahead to develop partnerships, see Section 10, 'Creating program capacity for cluster detection and response.'

The role of a communication plan in interventions

- Messaging with the affected population, providers, and the general public can be an important component of an intervention strategy. Planning for a communication strategy should ideally start before any cluster is identified. For considerations in developing a communication plan, see Section 10, 'Creating program capacity for cluster detection and response.'

Protecting confidentiality

- Intervention approaches can raise concerns for privacy and confidentiality. Please see Section 8, 'Protecting cluster data,' for considerations regarding privacy and confidentiality in interventions.

Section 7. Monitoring progress

Monitoring the progress of interventions in transmission clusters is important, in order to 1) understand whether transmission has been successfully interrupted or whether more work is needed, 2) to assess the impact of the investigation and interventions, and 3) to provide information to improve future investigation and intervention activities. As we gain more experience with this new activity, this type of information will provide information about the tangible outcomes of these investigations and interventions, help to guide priorities in future investigations, and provide lessons learned.

- **Has transmission been interrupted, or is it ongoing?**
    - Monitoring transmission clusters over time to identify new cases added to the cluster is the key strategy to understanding if transmission has been successfully interrupted. Sequence data should be analyzed routinely to identify any new cases in a cluster.
        - Keep in mind that the detection of new cases in a cluster will be very sensitive to the completeness of sequence data. To address this, it is important to understand what populations are not represented in sequence data, and identify areas where there are opportunities to work with providers to ensure that they are conducting resistance testing, and to work with labs to ensure that they are submitting sequences.
        - When new cases are identified, include them in the cluster investigation to determine if there is new information that increases understanding of factors facilitating transmission or provides a refined understanding of opportunities for intervention.
    - Additional cases in a cluster might not reflect continued transmission, as case finding activities resulting from the investigation could result in newly diagnosed cases that had already acquired HIV before the time of the investigation. Continued monitoring over

time can reveal unexpected new diagnoses after case finding activities have ceased, which could suggest that transmission has not been interrupted.

- Consider how new HIV-positive persons in the network are being identified. Are newly identified HIV positive persons in the network being detected result of active case finding activities, or through passive means, such as presentation for symptoms or through screening such as for plasma donation?

- Consider the potential for future growth, based on characteristics of persons identified in the transmission/risk network. Do HIV-infected persons in the network remain virally unsuppressed, or have previously suppressed persons become unsuppressed? Are there other factors that raise concern for ongoing growth, such as ongoing risk behaviors, or evidence that a network was not adequately captured through investigation and intervention? If any of these situations is the case, consider the potential for ongoing intervention activities.

- **What was the impact of the investigation, and of subsequent interventions?**
  - To assess the impact of the investigation, systematically gather information on what was learned from the investigation, and what actions took place as a consequence of the cluster investigation. Data collected should include:
    - Had the cluster been identified through other means (e.g., partner services), prior to detection through sequence analysis?
    - What was the full size of the underlying transmission/risk network identified through cluster investigation?
      1. Number identified as part of the molecular cluster
      2. Number of HIV-positive persons without sequence data linked to the cluster
      3. Number of HIV-negative persons in network at risk of infection.
    - What individual-level interventions were implemented for persons in the transmission cluster and risk network? Data collected should include:
      1. HIV-infected persons identified as out of care who were linked to care
         a. Number achieving viral suppression
      2. Partner services interviews initiated
         a. Previously unidentified partners named as a result of interview or re-interview. Of these:
            i. Number tested
               1. New HIV-positive partners
               2. Previous HIV-positive partners
               3. HIV-negative partners
            ii. Not tested
               1. Out of jurisdiction
               2. Not located
               3. Located but not tested
      3. HIV-negative persons in the risk network referred for PrEP
    - Did venue-based testing or other case-finding activities take place? If so, collect information on:
      1. Number of persons tested
         a. Number of newly diagnosed HIV-positive persons identified
         b. Number of previously diagnosed HIV-positive persons identified
         c. Number testing HIV-negative
            i. Of these, number evaluated and referred for PrEP

        - Did changes in policy or protocol result from the investigation, such as changes to the process for partner services interviews? If so, collect information on what actions were taken.
        - Did education or communication efforts take place as a result of the investigation? This could include provider or community outreach, media campaigns, or heath alert notifications. If so, collect information on what actions were taken.
    - Impact can also be considered in the context of the likely number of infections averted as a result of an investigation and subsequent intervention activities. Such an assessment would likely come from model-based estimations. Approaches for these types of estimates have not yet been developed.
- **What are lessons learned for future cluster investigation and intervention activities?**
    - Following a cluster investigation and intervention activities, consider conducting a 'hot wash' or other systematic assessment of what activities worked well and what activities were less effective. This can help inform strategies for future cluster investigations and interventions, and help ensure that resources are utilized as effectively as possible. Questions to consider include:
        - Which components of the cluster investigation yielded the most useful information?
        - Which data sources were most useful?
        - What were the staffing/resource needs for the investigation and intervention activities?
        - Which partnerships were most effective, and which could benefit from additional development?
        - What were the costs associated with cluster investigation? With intervention?

Section 8. Protecting cluster data

Protecting cluster data

Data regarding clusters of concern is inherently sensitive, and should be handled in accordance to state/local Public Health Law and procedures outlined in the NCHHSTP Data Security and Confidentiality Guidelines, including, but not limited to, the following:

- All data should be stored securely whether in electronic or paper form
- Access to identifiable information should be limited to authorized persons
- Any electronic output that could breach confidentiality (e.g., line listings, STATENOs) should be stored on a secure server. Hard copies should only be produced when necessary; when produced, they should be locked up and not taken out of the office. Paper copies should be shredded when no longer in use.
- All confidential data (including line lists and any documents including STATENOs) should be marked as confidential and encrypted for transfer or when not in use.
- Any information taken into the field as part of field investigation or service provision should include only the minimum amount of information necessary and be maintained securely at all times. Areas should develop specific procedures for securing information during field investigations.
- Data should only be shared with staff who have a need to know the information
- Any breach of data security protocol, regardless of whether personal information was released, should be reported to the overall responsible party and investigated immediately.

Additionally, ensure data on clusters of concern are handled consistently with any state or local Public Health Law and privacy and confidentiality guidelines relevant to this activity.

Because cluster data can be used to infer potential transmission linkages between persons, there are additional sensitivities in using this data beyond those inherent to HIV surveillance data. Particular considerations include:

- Data on potential transmission linkages between persons (i.e., which pairs of people have genetically linked sequences) may be subject to misinterpretation by those not familiar with this type of analysis. Consider minimizing use of these data and instead focusing on cluster-level data (i.e., considering all people in a cluster for intervention rather than focusing on people based on their position in the cluster)
- Communication about transmission clusters should be considered carefully to ensure that the privacy and confidentiality of members of a cluster is protected. Particular consideration should be made when communicating about a cluster with:
    - Frontline staff (e.g., DIS who might re-interview cluster members)
    - Healthcare providers
    - Persons involved in the transmission cluster and risk network (e.g., when interviewing or re-interviewing persons)
    - Community-based organizations and other potential referral resources
    - The general public (e.g., press releases or other messaging about a transmission cluster)
- For each of these audiences, consider that minimal amount of information about the cluster needs to be shared in order to communicate effectively, while not sharing more than is needed.
- Some jurisdictions may have laws that criminalize HIV transmission. Programs should work with general counsel to evaluate the implications of such laws and criminal statutes in their jurisdiction and ensure transmission cluster data are adequately protected.
- Information about transmission clusters should be used only for public health purposes. Programs should have discussions about what protections exist and whether they are adequate to ensure that information will not be released for non-public health purposes.

Section 9. Approaching multijurisdictional clusters

What is a multijurisdictional cluster?

- Multijurisdictional clusters are those in which coordination across jurisdictions is required for effective investigation and response. Often, multijurisdictional clusters will include cases reported from multiple states, however the jurisdictional issues involved could be relevant for clusters involving multiple counties within a single state, particularly if they include separately funded HIV surveillance or prevention programs. Some considerations for multijurisdictional clusters might also be relevant for clusters involving institutions such as correctional or military facilities, where coordination and data sharing across systems would be required.

Identifying clusters with multijurisdictional involvement

- Clusters that involve cases reported by multiple states can only be identified through national analyses conducted by CDC. State and local-level analyses might also identify clusters that involve multiple states, however, if persons in the cluster have moved to another state since diagnosis. Additionally, state-level analyses can identify clusters that span multiple counties.

Communicating with other jurisdictions involved

- Who takes the lead on an investigation involving multiple jurisdictions?
    - Generally, the jurisdiction with the majority of cases in a cluster, referred to by CDC as the ‘primary jurisdiction’, will take the lead on a cluster investigation. In situations where multiple jurisdictions are involved and no single jurisdiction has a majority of cases, CDC may take the lead, organizing data collection, investigation, and intervention efforts.

    - The lead jurisdiction has the responsibility for coordinating and leading the investigation, and keeping all involved jurisdictions informed of investigation progress and findings.
    - All involved jurisdictions involved should be involved in synthesizing investigation findings across multiple jurisdictions, and making decisions about next steps and interventions that will have implications for the jurisdictions involved.
- Why should programs reach out to other jurisdictions about cases in a cluster?
    - When a cluster involving multiple jurisdictions is identified, communication across jurisdictions will be important for effective cluster investigation and intervention. This is because a comprehensive case definition that includes all identified cases in a cluster is key to understanding the extent of transmission in a transmission cluster, to identify characteristics of cases and factors facilitating transmission, and to determine opportunities for effective intervention.
- How should programs communicate with other jurisdictions about a cluster?
    - The most effective approach to communicate with another jurisdiction about a cluster is to reach out directly to the surveillance coordinator and Molecular HIV Surveillance coordinator (if applicable) for that jurisdiction. For multijurisdictional clusters identified by CDC, CDC facilitates communication by sharing limited cluster information with surveillance and Molecular HIV Surveillance coordinators for all jurisdictions involved.
    - For jurisdictions that do not participate in Molecular HIV Surveillance, additional background and context might be needed in order for the jurisdiction to understand what information is needed and why. In these situations, consider contacting CDC for assistance facilitating communication and providing any needed background information.
- What information should be requested?
    - When you reach out to another jurisdiction about a case involved in a cluster, consider requesting the same information that you are gathering from readily available data sources in your own jurisdiction. Such information could include data from surveillance (e.g., demographic, risk, and clinical information), information on care status from other data sources, and information from partner services, including information on named partners that could meet the cluster case definition. Additionally, assess whether any connections between the jurisdictions were identified.
    - As the investigation progresses, continued communication and coordination could be needed.
- What to do when you are contacted by another jurisdiction about a cluster
    - When contacted by another jurisdiction for information on a case involved in a cluster, please prioritize sharing information to the extent allowed by your state and local data sharing policies. Even if few or just one case in your jurisdiction is identified by sequence data as belonging to a transmission cluster, the true extent of the transmission cluster in your jurisdiction could be much larger, and investigation and intervention efforts could present prevention opportunities that could reduce transmission in your jurisdiction as well as the primary jurisdiction.
- Special considerations for communication and coordination across jurisdictions
    - Jurisdictions may differ in the relative priority given to a particular cluster investigation, and relative resources available. Consider if higher level leadership discussions might be needed, to determine joint priorities and available resources.
    - Jurisdictions might also differ on legal restrictions in data sharing. Developing plans for data sharing early, and developing data sharing agreements if needed, can help facilitate communication and coordination.

- Effective communication and coordination could require the development of new collaborations/relationships between jurisdictions, or the strengthening of already existing relationships. Such relationships will provide a strong foundation for future cluster investigations.
- For jurisdictions within a state, existing contracts (such as those from a state funding county programs) could have implications for the scope of work a given jurisdiction can take on in the context of a cluster investigation.
- The needs for communication and coordination across jurisdictions could evolve as an investigation expands.

Section 10. Creating program capacity for cluster detection and response

Health departments should work to create a program with capacity to identify, review, prioritize, investigate, and intervene in clusters, ideally before any clusters are identified. Having a clear process in place will help facilitate action when clusters are identified. The key components of this work include:

Identify key staff and establish roles

- Identify person(s) who will be responsible for routine analysis of data to identify clusters. Ensure that they have the needed training and resources to conduct this activity. Also consider resources that might be available when surge capacity is needed, such as outbreak response staff outside of the HIV program.
- Identify key personnel and communication processes for cluster review, assessment, prioritization, decision making, and related resource-allocation. This group should review information for clusters identified by CDC and those identified locally.
    - Note that this should involve staff from both surveillance and prevention programs; because prevention programs will typically be the ones to implement interventions, their participation in these processes is critical. Partner services staff, who will typically be an important contributor to these discussions, may fall under STD or HIV prevention programs. This review should be conducted by staff with epidemiology expertise.
- Identify key stakeholders and partners who may need to be informed or involved in a cluster investigation or response, depending on the situation. Key stakeholders and partners who may need to be informed or involved include (but are not limited to):
    - Health department leadership and staff
    - Surveillance and prevention program leadership
    - Health department media/communication contacts
    - Community-based organizations
    - Community planning group
    - Ryan White HIV/AIDS Program grant recipients including service providers
    - Health department legal staff
    - Providers in key HIV care facilities
- Consider incorporating decision-making about situations in which a higher level response should be activated, and appropriate staff to be involved in these situations (e.g., public health emergency preparedness staff at the health department)

Determine how to identify clusters

- Currently, CDC conducts quarterly analyses of surveillance data to identify clusters meeting CDC priority criteria, which are then communicated to states

- Jurisdictions conducting their own analyses of sequence data locally, it will be important to identify the person(s) responsible for these analyses and ensure adequate training to conduct these analyses.
- Determine the frequency with which analyses will be conducted. More frequent analyses will allow more rapid detection of growing clusters and therefore greater opportunity for effective investigation and intervention. The advantages of more frequent analysis should be balanced with staffing and resource availability to determine the frequency for a given program.
- For more information on strategies to identify growing transmission clusters, refer to Section 3, 'Identifying growing transmission clusters.'

Establish key criteria for cluster review, review frequency, and prioritization for investigation and public health action

- Determine priority criteria for clusters that will be undergo preliminary investigation. While CDC has standard criteria to identify priority clusters (current criterion is ≥5 cases in the most recent 12-month period in a cluster defined at a 0.5% threshold), programs might want to consider alternate criteria. Such criteria could include lower thresholds for recent growth (e.g., 3 or 2 newly diagnosed cases in the most recent 12 month period), or criteria that focuses on particular risk groups (e.g., a demographic group with recent increases in diagnoses). Such criteria should be considered in the context of the resources available for review and possible investigation of clusters, as less stringent criteria or multiple priority groups could result in a larger number of clusters identified.
    - Determining priority criteria may be an iterative process that evolves over the course of months or years as a program gains experience with cluster detection and response.
    - Other cluster detection approaches, such as those based on space/time clustering of diagnoses in the absence of molecular data, might also be considered in a prioritization scheme.
    - Once a cluster has been identified as a priority cluster, it may be useful to determine whether any persons with HIV diagnosed more than 3 years ago are linked to the cluster, as these persons could be not virally suppressed and contributing to transmission
- Identify readily available data for 'desk review' of clusters, and develop processes for accessing and linking that data to cluster information.
- Develop a process to re-review and prioritize clusters when new information becomes available

Assess local policies that might impact cluster investigation and potential intervention

- Programs might have policies in place that present potential obstacles to cluster investigation or potential intervention, such as policies limiting re-interview of persons by partner services. Review and assess these policies to identify any potential obstacles, and consider if there are alternative approaches, or opportunities to revise policies to accommodate cluster investigation and intervention activities.
- Consider incorporating these policies into existing outbreak response plans or other response plans already developed by the health department that could provide guidance for available resources and strategies in the context of a potential HIV cluster investigation and response.
- Also consider the importance of protecting privacy and confidentiality during a cluster investigation. Please see Section 8, 'Protecting cluster data,' for considerations regarding privacy and confidentiality in cluster investigations.

Identify potential local resources that could be mobilized for investigation/response if needed

- Cluster investigations may require a larger number of staff for short periods of time. Consider staff who could be available for investigation activities. In jurisdictions with many HIV diagnoses, cluster investigation activities may be ongoing, necessitating staff whose time is fully allocated to this work. However, in jurisdictions with fewer diagnoses, staff may not to be dedicated to this activity full time, but rather would be pulled from routine work for cluster investigation

activities when needed. Such staff should ideally form a multidisciplinary team, with the range of skillsets needed for an investigation (e.g., data management and analysis, knowledge of partner services, familiarity with medical chart review, and interview).

- Identify local resources and programs that could be mobilized as partners for a potential intervention if needed. Such resources could include referral sites, testing and vaccination programs, and community-based organizations that routinely work with particular populations.
    - Consider how these resources could be mobilized if needed. Preparation could include standing up a multidisciplinary workgroup to broker resources and developing policies/procedures to facilitate coordination
    - Consider the importance of utilizing existing mechanisms/structures and interventions where available, rather than developing a new mechanism or structure.
- Develop relationships with community-based organizations and other partners for potential intervention activities
    - Create memoranda of understanding that might be needed in advance of any collaboration, to facilitate rapid action when timely response is needed.

Create a communication plan

- Identify key internal and external stakeholders who may need to be informed of a cluster investigation. Such stakeholders could include (but are not limited to):
    - Health department leadership
    - Surveillance and prevention program leadership and staff
    - DIS and front-line staff
    - Health department media/communication contacts
    - Health department legal counsel
    - Local health department staff and leadership
    - Community-based organizations
    - Ryan White HIV/AIDS Program grant recipients including service providers' care facilities leadership or leadership from other key care facilities
    - Correctional facilities, if applicable
    - Military facilities, if applicable
    - Tribal organizations, if applicable
    - Behavioral health providers, if appropriate
- Consider potential approaches to communicate with affected persons or the larger at-risk community.
- Identify opportunities to communicate with providers, such as health alerts. Consider potential messaging for such alerts and identify examples that can be drawn from. Note that STD and partner services staff might have more experience developing these alerts.
- Prepare for the potential to communicate through the media
    - Consider developing press release templates, even if not needed initially. In the context of an investigation or response, a program might want to have a press release developed even if it is not proactively released. Prepare for the potential for multiple media requests.
    - Identify the Health department press officer who would be the first media contact. Consider early education opportunities for this staff person to make them aware of cluster detection and investigation activities.
- Develop a plan to maintain regular communication with stakeholders throughout an investigation and response activities. Such a plan could include routine (e.g., weekly) updates.
- Consider the potential that messaging could stigmatize affected populations; craft messaging approaches carefully to avoid creating or contributing to stigma.

Appendix A: List of abbreviations and key definitions

AIDS Drug Assistance Program (ADAP)	State and territory-administered program authorized under Part B of the Ryan White HIV/AIDS Treatment Extension Act of 2009 that provides FDA-approved medications to low-income people living with HIV who have limited or no health coverage from private insurance, Medicaid, or Medicare. ADAP funds may also be used to purchase health insurance for eligible clients and for services that enhance access to, adherence to, and monitoring of drug treatments *Source: Health Resources & Services Administration*
Community-based organization (CBO)	Provides HIV prevention services, including HIV counseling and testing, to populations that are hard to reach and at high risk for transmitting or acquiring HIV. CBOs can act as a partner services entry point for clients who might not otherwise be offered these services, and staff members can promote partner services to the communities. CBOs also might be adept at gaining trust and establishing rapport with wary, untrusting, and fearful clients and their partners. CBO staff members might effectively elicit partner information from HIV-infected clients and provide counseling and testing to partners who come to the CBOs for services. Before partner services program managers determine how best to use CBOs in the partner services process, they should consider local laws and regulations. In certain jurisdictions, health departments and medical providers are the only entities with legal authority to notify persons of their exposure to HIV and other types of STDs.
Cluster investigation	A systematic process to: 1) Identify the underlying transmission cluster and risk network (e.g., undiagnosed cases, diagnosed cases without a sequence, persons at risk for HIV) 2) Identify factors possibly associated with transmission 3) Understand possible connections between cases in a cluster 4) Assess potential risk for ongoing transmission 5) Determine what potential interventions might be effective
Cluster snapshot	A document developed by the CDC HIV Incidence and Case Surveillance Branch to communicate cluster and case-level data on a molecular cluster to state and local health departments. An example of a cluster snapshot can be found in Appendix E.
Disease intervention specialists (DIS)	Health department personnel who are specifically trained to provide partner services. Some health departments use different titles for persons providing partner services. In addition, in certain jurisdictions, other persons (e.g., HIV counselors or clinicians) either inside or outside of the health department provide certain or all elements of partner services.
Drug resistance testing	Conducted in order to identify mutations associated with viral resistance to antiretroviral medications and help the HIV care provider select an appropriate treatment regimen. Drug resistance testing is recommended for all persons with diagnosed HIV infection, with the recommendation that testing be conducted at

	entry to HIV care. Drug resistance testing is typically ordered by providers at entry to HIV care, but can also be ordered at later time points (for example, if a person is on treatment but does not have a suppressed viral load). A nucleotide sequence is generated as an intermediate byproduct from a drug resistance test.
Engagement in care	Measured by whether a person with diagnosed HIV infection has a had at least one HIV medical care visit during the analysis period
Genetic distance threshold	The level of genetic similarity used to identify closely related pairs of sequences. The genetic distance threshold used can vary based on the goal of the analysis.
HIV TRAnsmission Cluster Engine (HIV-TRACE)	A bioinformatics tool developed by researchers at the University of California, San Diego to analyze nucleotide sequences and identify clusters representing recent and rapid transmission. A secure local installation of HIV-TRACE at CDC is used to run routine analyses on national surveillance datasets.
Molecular cluster	Identified through analysis of HIV genetic sequence data that is generated through HIV drug resistance testing. Molecular clusters contain only those people for whom molecular data is available and can be analyzed, and contains a subset of what is likely a larger underlying transmission cluster
Molecular HIV Surveillance (MHS)	A component of the National HIV Surveillance System. CDC funds selected state and local health departments to conduct molecular HIV surveillance activities.
Multijurisdictional cluster	A cluster in which coordination across jurisdictions is required for effective investigation and response. Often, multijurisdictional clusters will include cases reported from multiple states, however the jurisdictional issues involved could be relevant for clusters involving multiple counties within a single state, particularly if they include separately funded HIV surveillance or prevention programs.
National HIV Surveillance System (NHSS)	The primary source for monitoring HIV trends in the United States. The primary functions of the National HIV Surveillance System (NHSS) are (1) to provide accurate epidemiologic data to monitor the incidence and prevalence of HIV infection and HIV-related morbidity and mortality and (2) to use these data trends to assist in public health planning and policy. CDC provides federal funding to states and territories through surveillance cooperative agreements to achieve the goals of NHSS and also to assist states in developing their own surveillance programs in accordance with state and local laws and practices.
Nucleotide sequence	An intermediate byproduct of an HIV drug resistance test. Analysis of nucleotide sequences can identify pairs of sequences that are extremely similar and which may be closely related in transmission
Outbreak	An increase, often sudden, in the number of cases of a disease above what is normally expected in that population in that area.
Partner services	A broad array of services that should be offered to persons with HIV infection, syphilis, gonorrhea, or chlamydial infection and their partners. A critical function of partner services is partner notification, a process through which infected persons are

	interviewed to elicit information about their partners, who can then be confidentially notified of their possible exposure or potential risk. Other functions of partner services include prevention counseling, testing for HIV and other types of STDs (not necessarily limited to syphilis, gonorrhea, and chlamydial infection), hepatitis screening and vaccination, treatment or linkage to medical care, linkage or referral to other prevention services, and linkage or referral to other services (e.g., reproductive health services, prenatal care, substance abuse treatment, social support, housing assistance, legal services, and mental health services).
Pre-exposure prophylaxis (PrEP)	A way for people who do not have HIV but who are at substantial risk of getting it to prevent HIV infection by taking a pill every day
Primary jurisdiction	The jurisdiction with the majority of cases in a molecular cluster
Priority cluster	A molecular cluster that has met certain criteria and which should be flagged for preliminary investigation. Currently, CDC-defined priority clusters are clusters identified at a 0.5% genetic distance threshold with ≥5 cases in the most recent 12-month period.
Ryan White HIV/AIDS Program	Provides a comprehensive system of care that includes primary medical care and essential support services for people living with HIV who are uninsured or underinsured. *Source: Health Resources & Services Administration*
Secure HIV-TRACE	A web-based bioinformatics tool developed by researchers at the University of California, San Diego and Temple University to analyze HIV nucleotide sequences and identify molecular clusters. Secure HIV-TRACE is available to individual public health institutions to facilitate real-time analysis by state and local health departments to better understand and respond to their specific HIV burden.
Social network testing strategy	A recruitment approach for reaching and providing HIV counseling, testing, and referral services to persons who are unaware of their HIV infection by using existing social connections
Transmission cluster	A group of HIV-infected persons (diagnosed and undiagnosed) who have a direct or indirect epidemiological connection related to HIV transmission. A transmission cluster represents a subset of an underlying risk network
(Underlying) risk network	Includes the group of persons among which HIV transmission has occurred and could be ongoing. This network includes persons who are not HIV-infected but may be at risk for infection, as well as the HIV-infected persons in the transmission cluster

Appendix B. Frequently asked questions about HIV-TRACE and transmission network analysis
Excerpted from 'Secure HIV-TRACE: a guide for public health departments to reconstructing HIV-1 transmission clusters,' courtesy of Joel Wertheim.

Why pairwise alignment?
Secure HIV-TRACE was designed to detect transmission clusters using the 1497 nucleotide region spanning the HIV-1 *pro/rt* region common in public health surveillance activities, drug resistance screening, and research studies. This genomic region is from a conserved genomic region with very limited length variation (unlike, say, *env*) across all major HIV-1 subtypes and circulating recombinant forms. The absence of insertions and deletions permits robust pairwise alignment to a reference sequence. This approach is a timesaving measure compared with the more computational intensive approach of multiple sequence alignment, because it has linear complexity in the number of sequences; popular multiple sequence alignment algorithms all have superlinear complexity. **Secure HIV-TRACE** uses a modified version of the Smith-Waterman algorithm, which aligns nucleotide sequences by considering amino-acid translations of constituent codons; this approach allows us to make full use of amino-acid conservation to preserve alignment accuracy for divergent sequences (e.g., those from different subtypes).

Why genetic distance?
Genetic distance can be used as a proxy for epidemiological relatedness, because it increases as a function of time since transmission (in a linear fashion, as a first order approximation). This increase in genetic distance, due to an underlying molecular clock provides us with a proxy for the amount of time that has passed since two viral strains diverged from one another. The molecular clock in HIV, however, is highly imprecise because of factors like latency and natural selection due to immune escape and anti-retroviral treatment. Furthermore, the virus evolves in both the donor and recipient, so the distance between two strains is not simply a multiplier for the time since transmission. That being said, genetic distance serves as a useful proxy for epidemiological relatedness.

Why use a fixed distance cutoff?
HIV *pro/rt* diverges from the founder strain at a rate of about 0.1% per site per year, which indicates that a cutoff of 0.015 is about 5-15 years of combined evolution (in the source and the recipient). Distributions of pairwise distances in large sets of sequences (i.e., local, national, regional, and global) have the characteristic property of resembling a mixture of two distributions (see **FIGURE B1**): a component near 0 (i.e., closely/recently related sequences) and a component near 0.06 (i.e., two random sequences of the same subtype). Distance cutoffs of 0.01 to 0.02 segregate the two components nicely. See more below: **How do I select a genetic distance threshold?**). Further, our recent work in named partners in New York City has demonstrated that genetic distance alone provides better insight into who are potential transmission partners than partner tracing alone. In a sense, using genetic distances allows one to perform something analogous to contact tracing among all persons in a surveillance cohort, asking each pair if they have an epidemiological connection.

What is TN93 genetic distance?
TN93 is the name of a nucleotide substitution model developed by Koichiro Tamura and Masatoshi Nei, published in 1993. Hence, TN93. Nucleotide substitution models are used in evolutionary analyses to correct for multiple substitutions and/or reversions at a given site. Highly divergent sequences, with a greater number of substitutions separating them, are more likely to require complicated evolutionary models to properly estimate the level of divergence. The simplest evolutionary model, JC69, has a single parameter governing mutation rates among different nucleotides, and assumes equal frequencies for all nucleotides. In contrast, a more complex evolutionary model like general time reversible model with gamma rate variation (GTR+□) allows all nucleotide substitutions to occur at a unique rate, unique equilibrium base frequencies, and rate variation across sites. Importantly, over relatively short

evolutionary distances (i.e., <0.05 substitutions/site), GTR+Γ does not improve distance estimation accuracy for simpler models like JC69 (CITE), because not enough time has elapsed for a substantial number of multiple substitutions and/or reversions. In basic calculus terms, most curves resemble straight lines if you zoom in closely enough.

For **Secure HIV-TRACE**, we wanted an evolutionary model that optimizes both realism and computational efficiency. Simple models like JC69 and K2P (Kimura 2-parameter) have obvious shortcomings when applied to HIV: these models do not permit unequal nucleotide base frequencies, and HIV has notorious high frequencies of adenine (A) and low frequencies of uracil/thymine (U/T). The TN93 substitution model allows for unequal base frequencies and three different rates of substitutions between nucleotide bases: transitions between purines (i.e., A and G), transitions between pyrimidines (i.e., C and U/T), and transversions between purines and pyrimidines (e.g., A or U/T to C or G). Furthermore, distances estimated under TN93 can be represented by a closed form solution (i.e., computed without numerical optimization, simply from pairwise differences in nucleotide counts), which permits rapid computation of pairwise distances. More complex models require relatively expensive numerical optimization, especially because it will have to be done hundreds of millions or billions of times, to find all relevant distances. Therefore, when using genetic distances to identify potential transmission partners, which are expected to be between 0.01 and 0.02 substitutions/site divergent, a substitution model more complicated than TN93 is not needed, and there are no appreciable computational savings to be had by using cruder models. As an example, our implementation can compute approximately 10 million TN93 distances per second on a single server node.

Why not phylogenetics?

Phylogenetics is an extraordinary powerful tool for understanding viral evolutionary history and dynamics. That being said, its strength lies in its ability to say that two strains, Virus A and Virus B, are more closely related to each other (i.e., share a common ancestor more recently) than they are to a third strain, Virus C. This statement is relative and applies only to the viruses being considered. Moreover, this statement says nothing about whether the relatedness of Viruses A and B is epidemiologically meaningful. For example, any two subtype B sequences are more closely related to each other than either one is to a subtype D virus; to say that two randomly selected subtype B sequences have a meaningful epidemiological linkage, would be saying that we care about events that had happened more than 50 years ago. Although there have been studies that used only phylogenetic relatedness to establish HIV transmission clusters, our position is that just because something *was* done, does not mean it *should* have been done.

In fact, many HIV transmission network studies that used phylogenies **also needed a genetic distance** component: looking for groups of sequences that have low genetic divergence **and** high phylogenetic support (i.e., bootstrap, aLRT, or posterior probability). A bound on genetic distances establishes recency, whereas phylogenetics establish relatedness (relative to the rest of the sequences in the analysis). A major problem with relying on these phylogenetic support values to define what can be in a single cluster, is that they are highly contingent on the data, and change in counterintuitive ways. For example, as more sequences are added, bootstrap values can decrease, resulting in the breakdown of formerly robust transmission clusters. When the goal is tracking transmission network growth over time while adding more and more sequence data, this is a highly undesirable feature. Sequences that are clustered using **Secure HIV-TRACE** will always be clustered using **Secure HIV-TRACE**, if the analysis parameters stay the same. And adding more data can only increase the size of clusters, not break them apart.

Another issue with the phylogenetic approach is that it takes a lot of computational time, especially for big datasets with tens or hundreds of thousands of HIV sequences. Most of the time is spent determining the evolutionary relationships deep in the phylogenetic tree, which will never be considered in a study of

transmission clusters anyway. And when a few new sequences are added, the whole process needs to begin again. With our genetic distance approach, only the new sequences need to be considered, and all the previous computational work can be kept: like adding new pieces to a jigsaw puzzle.

Finally, while there exist approaches that estimate times along with trees (e.g., relaxed clock methods), they are so computationally expensive that they simply do not scale past about 1000 sequences, Moverover, they will typically give you essentially the same answer as **Secure HIV-TRACE**.

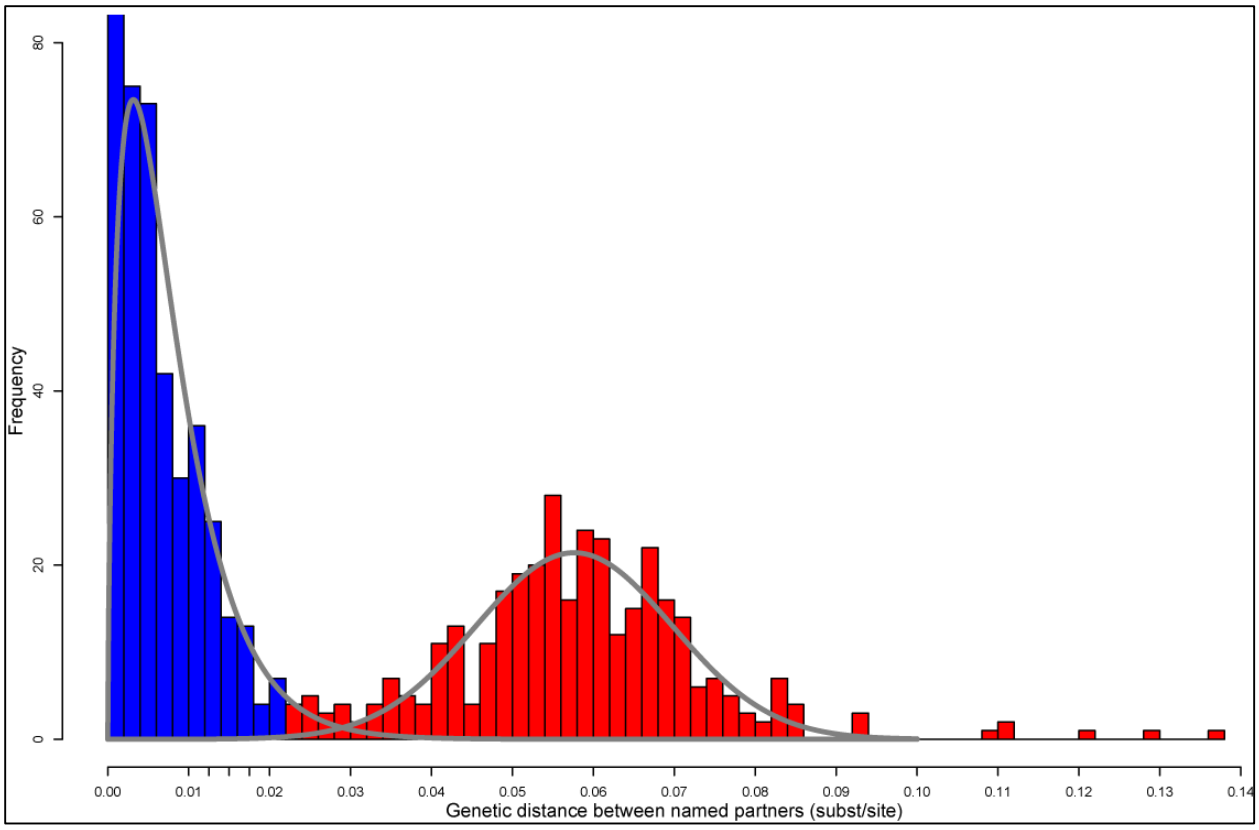


Figure B1. Distribution of genetic distances separating named partners in New York City. Potential transmission are shown in blue. Random within subtype variation is shown in red.

How do I select a genetic distance threshold?

An epidemiologically meaningful genetic distance threshold should link people who are potential transmission partners (i.e., close in the true transmission network) but not link people who are unlikely to have been involved in direct viral transmission. The best guide we have for determining a genetic distance threshold for identifying potential transmission partners in a U.S. surveillance setting comes from the analysis of 749 named partner pairs in New York City interviewed during 2006 through 2012. We analyzed the genetic distance separating baseline virus from named partners (reported sexual contact or shared injection drug use in the previous 12 months). When we plot these genetic distances, we observe two distinct modes: potential transmission partners (highlighted in blue) and partners who are HIV-infected, but have a genetic distance comparable to random within subtype variation (red). The potential transmission partners tend to have genetic distance ≤0.02 substitutions/site. To minimize the likelihood of spurious links in a surveillance cohort of thousands or tens of thousands of people, we recommend a slightly more conservative threshold: around 0.015 substitutions/site. More conservative genetic distance thresholds can also be applied to improve the probability that potential transmission partners share a meaningful epidemiological connection.

What are ambiguous nucleotides? Or ambiguities?

When HIV infects an individual, it forms genetically diverse and potentially complex populations within that person. Currently, the sequence data reported to the National HIV Surveillance System are produced using bulk Sanger sequencing, which produces a single genetic sequence representing this circulating population. This bulk sequence commonly reports diversity at polymorphic sites (i.e., when some strains have one nucleotide at a position and others have a different nucleotide at the same position) as common nucleotide IUPAC ambiguity codes [e.g., R (representing a mix of A and G), Y (a mix of C or T), N (a mix of all nucleotides)]. In standard phylogenetic inference, nucleotide ambiguities are "partially missing data" (e.g., Y is either C or T, but not A or G). When using pairwise distances (as in **Secure HIV-TRACE**) to construct genetic transmission networks, these nucleotide ambiguities have the potential to greatly complicate inference (see **FIGURE 25A**). The most conservative approach is to average the distance between ambiguities and resolved bases (e.g., Y is 0.5 differences from either C or T), and this is the approach we took when inferring the HIV-1 global transmission network. But averaging ambiguities in transmission network analysis penalizes sequences from chronically infected individuals—who are likely to have a more diverse viral population—and this averaging of distances makes these sequences less likely to cluster in the network. Therefore, resolving ambiguities (so that Y would be 0 differences from either C or T, and 1 difference from A or G) appears to be an attractive option. However, if we are too permissive in our tolerance of ambiguities, unrelated viruses can become connected in our network.

For example, if sequences from two people differ at 5% of sites, their viruses represent random intra-subtype variation and are not likely potential transmission partners. However, if within one of these people, most of this variation is polymorphic, and ambiguities are resolved in the genetic distance calculation, the genetic distance separating these viruses may fall below the distance threshold. Since variable sites are not uniformly distributed across the HIV genome, the highly polymorphic sequence is also likely to link to many other 'unrelated' viruses as well. The result is a large transmission cluster in which most sequences are connected to a hub (the high ambiguity sequence) but not to each other.

In an example from the San Diego Primary Infection Cohort (**FIGURE B2**), the genetic transmission network is affected by handling of nucleotide ambiguities. When ambiguities are fully resolved, the largest cluster in this cohort contains 119 people. However, when this cluster was mapped onto a maximum likelihood phylogenetic tree, its members are dispersed across the tree, encompassing the genetic diversity of the entire city of San Diego. Furthermore, the majority of nodes in the cluster are connected via two nodes acting as hubs (highlighted in red in **FIGURE B2a**) which have 5.8% and 7.6% ambiguities and represent the two highest degree nodes in the network. The nodes connected through the spokes on these hubs rarely share an edge with each other. This feature, along with the phylogenetic dispersion, suggests that this cluster is an artifact of nucleotide ambiguity resolution. When these two hubs are excluded from the analysis, the cluster breaks apart, resulting in several distinct clusters and unconnected nodes (**FIGURE B2b**).

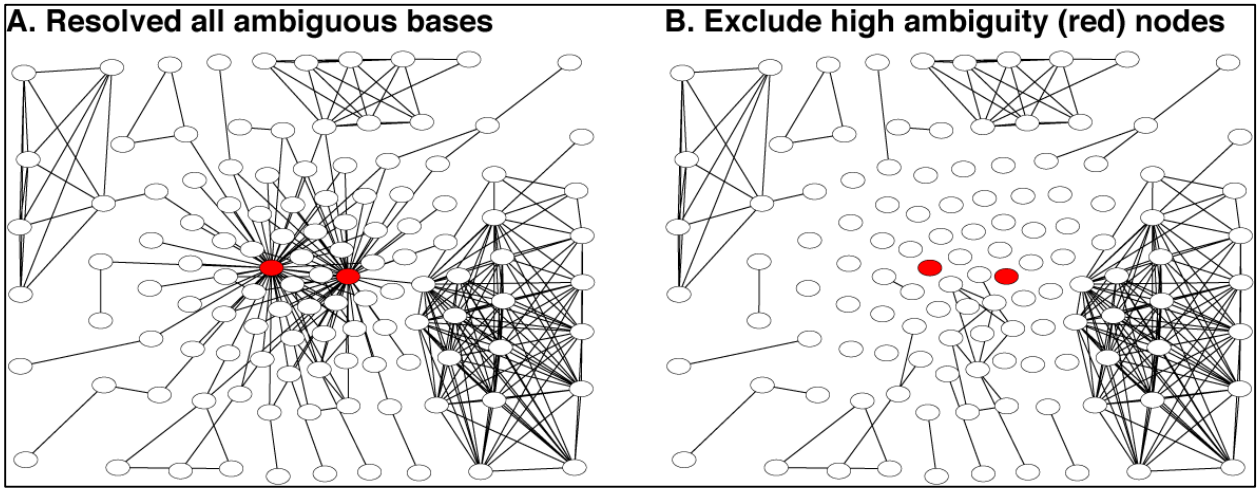


Figure B2. Example in which two contaminant sequences with high numbers of nucleotide ambiguities (shown in red) can create artificial clustering among unlinked singletons and unrelated clusters. (a) The inferred cluster resolving all ambiguous nucleotides. (b) The same cluster where the two hubs (shown in red) are excluded from the analysis.

Clusters that resemble **FIGURE B2** should be interpreted with extreme caution. They are almost always spurious and the result of erroneous inference due to high levels of nucleotide ambiguities (or contamination with "reference" strains).

How does Secure HIV-TRACE handle ambiguous bases?
We recommend that nucleotide ambiguities be fully resolved when calculating genetic distance only when (i) the sequences have a low proportion of ambiguities or (ii) if the size of the dataset is small. When constructing a transmission network for datasets of thousands or tens of thousands of sequences, we recommend penalizing sequences with high levels of ambiguities. The "Ambiguity Fraction" parameter governs this penalty. The default "Ambiguity Fraction" value of 0.015 resolves the genetic distance between ambiguous nucleotides when calculating the distance between sequences with ≤1.5% ambiguities and averages the genetic distance between ambiguous nucleotides when calculating the distance between sequences with >1.5% ambiguities.

Sequences with >5% ambiguous nucleotides will be flagged as problematic sequences and removed from the analysis. This protocol follows the guide set forth by the Los Alamos National Laboratory (LANL) HIV Sequence Database (https://www.hiv.lanl.gov/components/sequence/ HIV/search/help.html#bad_seq). Extremely high proportions of ambiguities can be the result of poor quality sequencing, contamination, or dual infection. Including these sequences can adversely affect the performance of **Secure HIV-TRACE**.

Why should I screen for laboratory contaminants?
Although the protocols for generating HIV-1 *pro/rt* genetic sequences are well validated, occasionally laboratory contamination with other genetic material is known to occur. This contamination is most often with the lab strain HXB2, but it can happen with any strain of HIV. Importantly, this contamination often results in a mixed sample where the resulting sequence is a combination of the isolate and the laboratory contaminant. This mixed sample often has high levels of ambiguous nucleotides and could compromise HIV-TRACE analysis if it were to be included, especially because mixing two unrelated strains will create ambiguities at many sites that tend to vary between strains, thereby enabling a "connection" through this sequence if ambiguous nucleotides are resolved (see above). Furthermore, if multiple contaminant sequences are included in the same analysis, they will erroneously be inferred to be part of the same

cluster. Therefore, we screen every run for HXB2 linked sequences. Any sequence that links to HXB2 is excluded from further analysis.

What about drug resistance associated mutations (DRAMs)?

DRAMs often arise in HIV found in people taking anti-retroviral therapy; they can be found in virus from both treatment-naive and treatment-experienced people who were initially infected with a drug-resistant virus. DRAMs typically occur at a select set of sites that are not polymorphic in the absence of prior anti-viral therapy. This type of convergent evolution at the amino acid-level has the potential to negatively affect phylogenetic inference [CITE]. The genetic distance separating two viruses that have undergone convergent evolution will theoretically be lower than two viruses that have not experienced convergent evolution. In practice, however, we find little to no effect of excising DRAM sites from network inference. Specifically, transmission networks built at the city, national, global level are robust to inclusion of DRAM sites. For example, when analyzing a cohort of named partner pairs in New York City, only a small fraction of partners become either linked or unlinked when DRAMs are excluded (red in **FIGURE B3**). Therefore, we do not recommend excising DRAMs from transmission network analyses using HIV-TRACE. An exception to this recommendation is for studies focusing on the effect of DRAMs on network characteristics; in these instances, DRAM site should be excised prior to network construction.

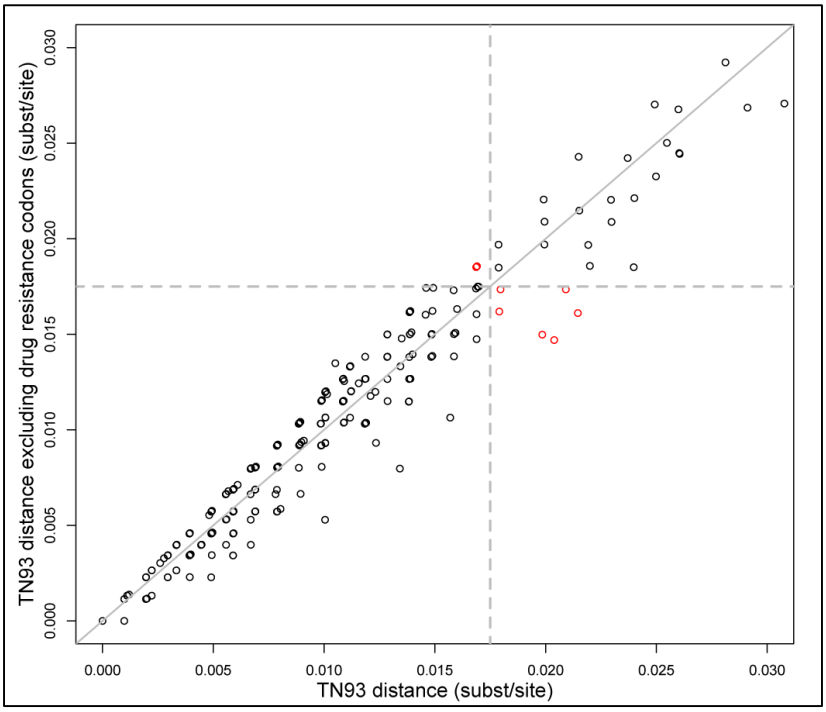
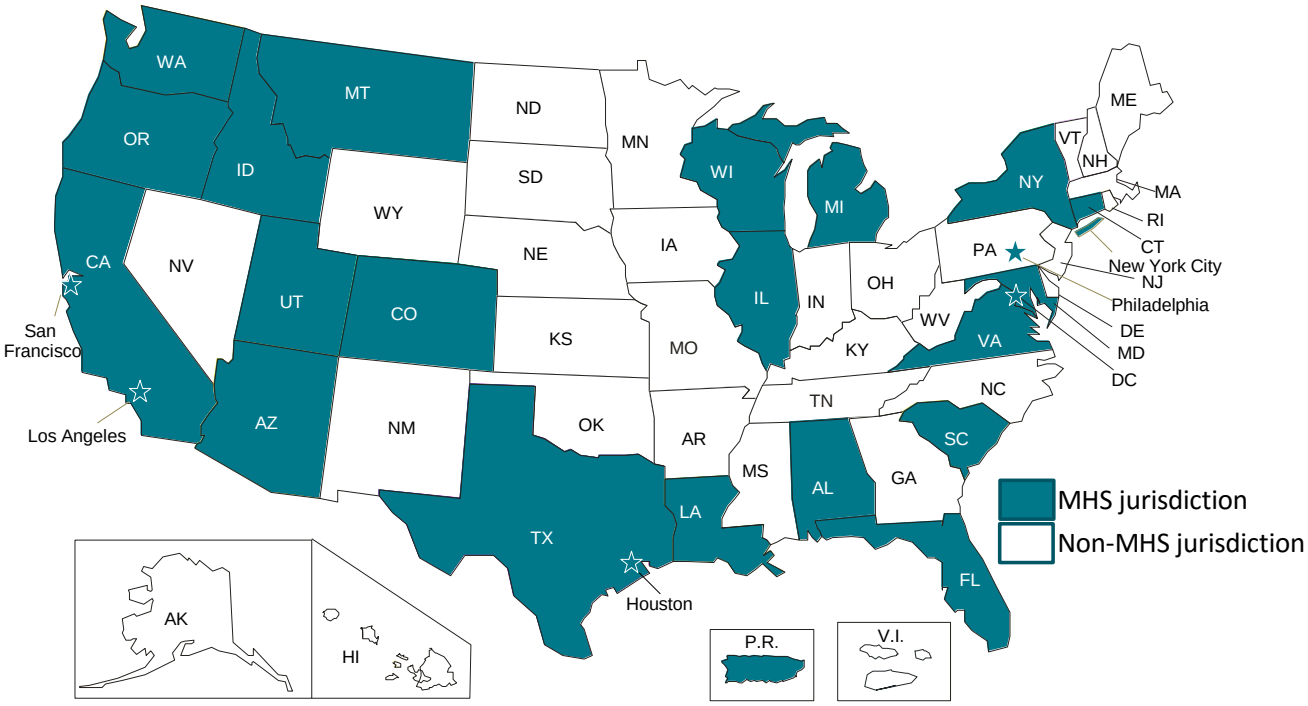


Figure B3. Genetic linkage including/excluding codons associated with drug resistance mutations in a New York City surveillance cohort. Nodes in red change linkage depending on inclusion/exclusion of DRAMs.

Appendix C. Map of Molecular HIV Surveillance jurisdictions



Appendix D. Additional resources

Example publications using HIV-TRACE:

- Wertheim JO, Leigh Brown AJ, Hepler NL, Mehta SR, Richman DD, Smith DM, Kosakovsky Pond SL (2014) The global transmission network of HIV-1 J Infect Dis 209(2): 304–313.
- Little SJ, Kosakovsky Pond SL, Anderson CM, Young JA, Wertheim JO, Mehta SR, May S, Smith DM (2014) Using HIV networks to inform real time prevention interventions PLOS ONE 9(6): e98443.
- Oster AM, Wertheim JO, Hernandez AL, Ocfemia MC, Saduvala N, Hall HI (2015) Using Molecular HIV Surveillance Data to Understand Transmission Between Subpopulations in the United States J Acquir Immune Defic Syndr 70(4):444–451.
- Whiteside YO, Song R, Wertheim JO, Oster AM (2015) Molecular analysis allows inference into HIV transmission among young men who have sex with men in the United States AIDS 29(18): 2517–2522.
- Wertheim JO, Oster AM, Hernandez AL, Saduvala N, Bañez Ocfemia MC, Hall HI (2016) The international dimension of the U.S. HIV transmission network and onward transmission of HIV recently imported into the United States. AIDS Res Hum Retroviruses.

Security and Confidentiality:

- Data Security and Confidentiality Guidelines for HIV, Viral Hepatitis, Sexually Transmitted Diseases, and Tuberculosis Programs (available at https://www.cdc.gov/nchhstp/programintegration/docs/PCSIDataSecurityGuidelines.pdf).

Social Network Testing Strategy:

- CDC. Use of social networks to identify persons with undiagnosed HIV infection—Seven U.S. cities, October 2003–September 2004. MMWR Weekly (2005). 54(24);601–605. Available at https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5424a3.htm
- McCree DH1, Millett G, Baytop C, Royal S, Ellen J, Halkitis PN, Kupprat SA, Gillen S. Lessons learned from use of social network strategy in HIV testing programs targeting African American men who have sex with men. AJPH 2013 Oct;103(10):1851-6.

Appendix E. Cluster snapshot companion document

HIV Cluster Snapshots: A Companion Document

HIV Incidence and Case Surveillance Branch
Division of HIV/AIDS Prevention
Centers for Disease Control and Prevention

February 2017

PURPOSE OF COMPANION DOCUMENT

This document serves as a companion guide to the HIV cluster snapshot generated by the HIV Incidence and Case Surveillance Branch (HICSB) at CDC. The snapshot provides summary information about a specific HIV molecular cluster that CDC has identified through analysis of the national HIV surveillance data, specifically HIV nucleotide sequences collected through molecular HIV surveillance activities.

The snapshot is generated for use by HIV surveillance programs to understand a specific HIV transmission cluster. Because the snapshot includes personally identifiable information, it should only be shared with health department colleagues on a 'need-to-know' basis and handled in accordance with NCHHSTP Data Security and Confidentiality Guidelines (www.cdc.gov/nchhstp/programintegration/docs/pcsidatasecurityguidelines.pdf).

KEY CONCEPTS

HIV Molecular Clusters

An HIV molecular cluster is defined as a group of persons with diagnosed HIV infection that have genetically similar HIV strains. The cluster represents a subset of an underlying risk network that can include persons with undiagnosed HIV infection, persons with diagnosed HIV infection who may or may not be in care, or HIV-negative persons at risk for acquiring HIV infection.

Cluster Identification

CDC has developed criteria to identify, using national HIV surveillance data, a subset of clusters with recent, rapid transmission that may require rapid and complete investigation and action. These clusters of concern are currently selected based on the following criteria:

1. Clustering of HIV nucleotide sequences at a low genetic threshold (0.5%), suggestive of recent transmission (pairs of sequences with very few genetic differences represent sequences that are very closely related)
2. At least 5 cases in the cluster diagnosed in the most recent 12 months included in the analysis, indicating rapid and recent growth

NOTE: A molecular cluster represents a subset of an underlying risk network in which transmission has occurred and could be ongoing. However, the molecular clusters cannot reveal which cases are directly related by transmission or determine the direction of transmission. This is because two persons with genetically similar HIV strains are not necessarily directly linked by transmission: the relationship could be indirect, and there could be unidentified persons involved in transmission relationships.

Information on Cases Reported by Other Jurisdictions

The amount of information provided in the snapshot on cases reported by other jurisdictions depends on whether the primary jurisdiction and the other jurisdiction have agreed, in accordance with their respective reporting and data sharing laws and regulations, to the reciprocal sharing of HIV surveillance data. When an agreement is in place, detailed case-level information for cases reported by other jurisdictions will be provided in the snapshot. Otherwise, only limited information will be shown.

PAGE 1 of SNAPSHOT

1

HIV TRANSMISSION CLUSTER SNAPSHOT: CLUSTER YYYYMM_####

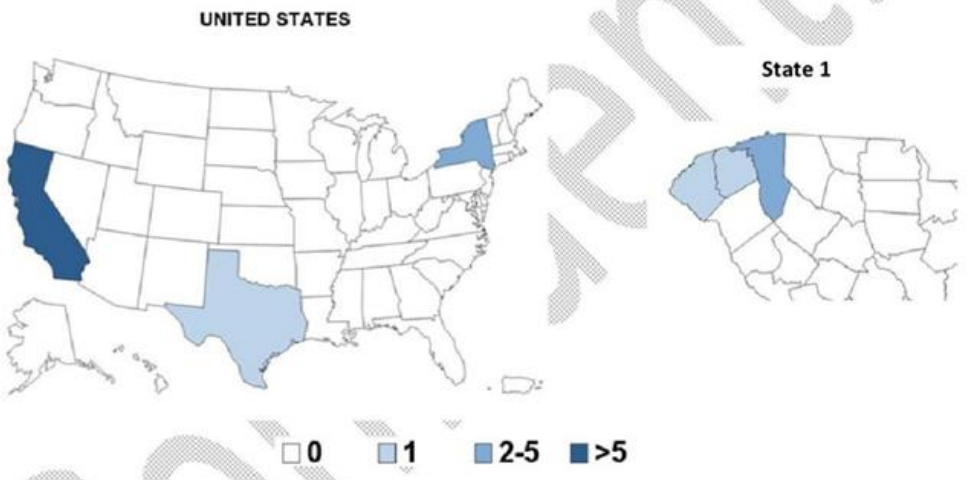
2 PRIMARY JURISDICTION: _____ 3 DIAGNOSIS DATES INCLUDED: MM/DD/YYYY–MM/DD/YYYY

Table 1: Overall information about analysis and data completeness.

4 **Dataset date:** Month YYYY	5 **Date of Analysis:** Month YYYY	6 **Threshold for cluster inclusion:** #.#%	
7 **% of diagnoses with HIV sequencing data available in primary jurisdiction, by year:**			
YYYY: ##% (##/##)	**YYYY**: ##% (##/##)	**YYYY**: ##% (##/##)	**YYYY**: ##% (##/##)
8 **% of diagnoses with HIV sequencing data available in all MHS jurisdictions, by year:**			
YYYY: ##%	**YYYY**: ##%	**YYYY**: ##%	**YYYY**: ##%
9 **Complete lab reporting of CD4+ and VL for primary jurisdiction during YYYY–YYYY?** Yes			
Total case count at #.#% threshold: 10 10			
Dataset in which cluster was first identified: MONTH YYYY 11			

Figure 1. Map(s) with number of cases by residence at diagnosis

12

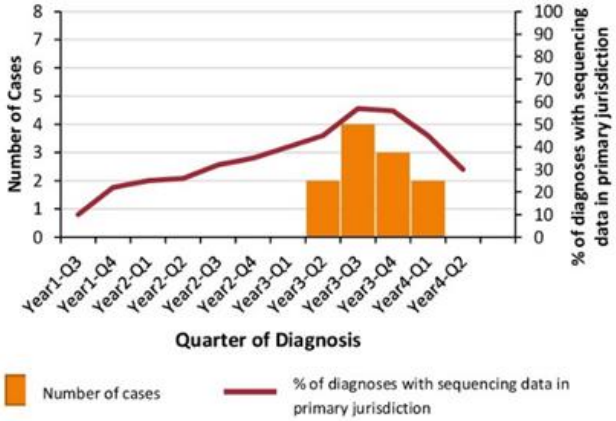


13 **Table 2. Residence at diagnosis.**

Residence at diagnosis	N = 11
STATE 1	N = 7
County 1	5 (78%)
County 2	1 (11%)
County 3	1 (11%)
STATE 2	N=3
STATE 3	N=1

Figure 2. Epidemiologic curve of cases in cluster and percent of diagnoses with sequencing data by quarter, Year1–Year4 (N=11).

14



DESCRIPTION of PAGE 1 of SNAPSHOT

Header:

1. The cluster number is a unique number assigned to identify each cluster. This number is composed of the year and month of the dataset in which the cluster was first identified as a priority cluster, followed by an arbitrary number.
2. The primary jurisdiction of residence at diagnosis is the jurisdiction with the majority of cases identified in the cluster using national data.
3. The cluster analysis includes data from the most recent three years of diagnoses, ending with the last day of the month of the dataset used for analysis. Limiting the analysis to the most recent three years assists in focusing efforts on identifying recent transmission.

Table 1:

4. This is the CDC dataset that was used to identify the cluster and generate the snapshot. All sequences ≥500 nucleotides in length collected for persons diagnosed in the years of interest were included. If a person had multiple sequences in one diagnosis year, the earliest and longest sequence was selected.
5. The month and year that the cluster analysis was conducted by HICSB.
6. A very low threshold (≤0.5% genetic distance) was used to identify extremely similar sequences. Although the calculation of genetic distance involves evolutionary models, the details of which are beyond this document, a genetic distance of 0.5% for two sequences that are each 1,000 base pairs long equates to a difference of approximately 5 nucleotides.
7. Completeness of sequence data for the primary jurisdiction, by year of diagnosis, is presented. The numerator is the number of cases with an eligible sequence for transmission network analysis, and the denominator is all cases residing in that jurisdiction at diagnosis. All eligible sequences, regardless of the time between the date of diagnosis and the date of specimen collection, are included. Because of reporting delay, completeness might be lower in later years compared to earlier years.
8. Completeness of sequence data for all MHS jurisdictions, by year of diagnosis, is presented for comparison.
9. Completeness of CD4+ and VL reporting for the primary jurisdiction is provided.
10. The total number of cases that clustered at 0.5% genetic threshold. The total number includes all cases in the cluster reported to the national HIV surveillance system by any jurisdiction.
11. The dataset in which the cluster was first identified.

Figure 1:

12. Jurisdictions with at least one case that clustered at the 0.5% genetic threshold are highlighted on a national map. A map highlighting the counties of the primary jurisdiction is also included in the snapshot.

Table 2:

13. The number and percent of cases identified in the cluster by county of residence at diagnosis for the primary jurisdiction are presented. These data correspond to the maps shown in Figure 1.

Figure 2:

14. The example epidemic curve displays the number of all cases (orange bars) that clustered at the 0.5% genetic threshold throughout the analysis, by year and quarter. The red line shows the completeness of sequence data by diagnosis year (same data shown in above) and is

independent of the number of identified cases indicated by the orange bars. Low sequence completeness suggests that the cluster could include additional cases not captured by sequence data.

PAGE 2 of SNAPSHOT

Table 3. Demographic, risk, and clinical characteristics of HIV cases in cluster

	Characteristic	Cases reported (N=7*)	
		n	Col %
15	**Age at HIV diagnosis (in years)**		
	<13	0	0
	13–19	3	43
	20–29	4	57
	30–39	0	0
	40–49	0	0
	50–59	0	0
16	≥60	0	0
	Sex		
	Male	7	100
17	Female	0	0
	Race/ethnicity		
	American Indian/Alaska Native	0	0
	Asian	0	0
	Black/African American	0	0
	Hispanic/Latino	0	0
	Native Hawaiian/Other Pacific Islander	0	0
	White	7	100
18	Other/Multiple race	0	0
	Transmission category		
	Male-to-male sexual contact (MSM)	7	100
	Injection drug use	0	0
	Heterosexual-Male	0	0
	Heterosexual-Female	0	0
	Male-to-male sexual contact and injection drug use	0	0
	Other/No identified risk	0	0
19	**Age of infection at diagnosis†**		
	Recent	0	0
	Longstanding	1	14
	AIDS within 6 months of HIV diagnosis	0	
	AIDS at HIV diagnosis	1	
	Other longstanding infection	0	
	Missing	6	86
20	**Evidence of viral suppression in past 12 months††**		
	Yes	3	43
	No	4	57
21	**Drug resistance**		
	Yes	2	29
	No	2	39
	Not determined	3	43

*Data reported by at least one other jurisdiction suppressed in this table, unless jurisdictions have agreed to share data

†Determined using STARHS results, when available, in combination with CD4-based staging data

†† Viral suppression defined as most recent viral load ≤200 copies/mL and collected ≤12 months before end of dataset (MMDDYYYY)

DESCRIPTION of PAGE 2 of SNAPSHOT

Table 3:

The demographic, risk, and clinical characteristics of the cases included in the cluster are shown in the table. These data are based on analysis of the national datasets. All cases reported exclusively by the primary jurisdiction will be included in the table. Cases that were not reported by the primary jurisdiction, or reported by both the primary jurisdiction and another jurisdiction, will only be included in the table if all reporting jurisdictions agreed to share data on the cases identified in the cluster.

15 The number and percent of cases by age group. Age is based on age at diagnosis.

16 The number and percent of cases by sex at birth.

17 The number and percent of cases by race/ethnicity.

18 The number and percent of cases by transmission category.

19 The number and percent of cases by age of infection at diagnosis. Age of infection is determined using STARHS (Serological Testing Algorithm for Recent HIV Seroconversion) results, when available, in combination with CD4-based staging data.

20 The number and percent of cases with any evidence of viral suppression in the past 12 months. Evidence of viral suppression was defined as most recent viral load ≤200 copies/mL and collected ≤12 months before the end of the dataset.

21 The number and percent of cases with any HIV drug resistance identified. Sierra, the Stanford HIV Web Service (Version 1.1), was used to assess the presence of mutations associated with 20 FDA-approved antiretroviral drugs that target the protease, reverse transcriptase and integrase enzymes of the HIV-1 *pol* region.

PAGE 3 of SNAPSHOT

Table 4. Line list (sorted by descending date of diagnosis)

Reporting jurisdictions ± (22)	StateNo (23)	Date of HIV diagnosis (24)	Facility at HIV diagnosis (25)	Residence at diagnosis (26)	Current residence (27)	Transmission category (28)	Sex (29)	Age (30)	Race / Ethnicity (31)	Recency (32)	AIDS at Diagnosis? (33)	Date of last negative* (34)	Most recent VL value (35)	Most recent VL date (36)	Whether drug resistance detected (37)	Number of mutations (38)	Specific drug resistance mutations (39)
STATE 1	######	4/4/YYYY	Internal Medicine	County 1, State 1	County 1, State 1	Male-to-male sexual contact	Male	20–29	White			2/YYYY	20	5/2/YYYY	Not determined	0	
STATE 1	######	5/21/YYYY	Internal Medicine	County 1, State 1	County 1, State 1	Male-to-male sexual contact	Male	13–19	White				120255	9/1//YYYY	Not determined	0	
STATE 1	######	7/7/YYYY	ER	County 1, State 1	County 1, State 1	Male-to-male sexual contact	Male	20–29	White			4/YYYY	30	4/15/YYYY	Not determined	0	
STATE 1	######	7/28/YYYY		County 1, State 1	County 1, State 1	Male-to-male sexual contact	Male	13–19	White			6/YYYY	1350025	10/12/YYYY	Yes	1	NNRTI_K103N
STATE 1	######	9/27/YYYY	ER	County 1, State 1	County 1, State 1	Male-to-male sexual contact	Male	13–19	White			5/1/YYYY	1151200	11/12/YYYY	Yes	1	NNRTI_K103N
STATE 1	######	9/21/YYYY		County 2, State 1	County 2, State 1	Male-to-male sexual contact	Male	20–49	White	Long Standing	AIDS at HIV diagnosis		20	6/18/YYYY	No	0	
STATE 1	######	10/14/YYYY	Univ Hosp	County 3, State 1	County 3, State 1	Male-to-male sexual contact	Male	20–29	White			3/18/YYYY	5320120	12/2/YYYY	No	0	
STATE 2	######			STATE 2	STATE 2												
STATE 2	######			STATE 2	STATE 2												
STATE 2	######			STATE 2	STATE 2												
STATE 3	######			STATE 3	STATE 3												

± List of all jurisdictions that have reported any data about this case

*Date of last negative HIV test is based on self-reported TTH information or a lab documented negative HIV test before HIV diagnosis. If dates from both sources are available, the more recent date is chosen.

DESCRIPTION of PAGE 3 of SNAPSHOT

Table 4:

The line list provides detailed information about each case identified in the cluster. These data are based on analysis of the national datasets. Detailed information for all cases reported exclusively by the primary jurisdiction will be included in the table. However, information on cases that were not exclusively reported by the primary jurisdiction will be limited and highlighted in grey at the bottom of the line list if the other jurisdictions do not agree to share data with the primary jurisdiction. The line list is sorted by descending date of diagnosis. NOTE: These data should be handled in accordance with security and confidentiality guidelines.

22 All jurisdictions that reported this case will be listed.

23 The StateNo assigned by the primary jurisdiction for each case will be displayed. StateNos assigned by other jurisdictions for the same case will be shared if that jurisdiction agreed to share data with the primary jurisdiction.

24 Date of HIV diagnosis was calculated based on national data. The date of HIV diagnosis will be shown for a case if all the reporting jurisdictions for a case have agreed to share data.

25 Facility at HIV diagnosis was based on national data. If all the reporting jurisdictions for a case have agreed to share data, then facility at HIV diagnosis will be shown.

26 Residence at diagnosis was based on national data. County and state at residence of diagnosis will be provided for cases reported by the primary jurisdiction and/or other jurisdiction(s) that agreed to share data with the primary jurisdiction. Otherwise, state at residence of diagnosis will be provided.

27 Current residence was based on national data. County and state of current residence will be provided for cases reported exclusively by the primary jurisdiction and/or other jurisdiction(s) that agreed to share data with the primary jurisdiction. Otherwise, state of current residence will be provided.

28 Transmission category was based on national data. If all the reporting jurisdictions for a case have agreed to share data, then transmission category will be shown.

29 Sex at birth was based on national data. If all the reporting jurisdictions for a case have agreed to share data, then sex at birth will be shown.

30 Age was based on national data. If all the reporting jurisdictions for a case have agreed to share data, then the age group will be shown.

31 Race/ethnicity was based on national data. If all the reporting jurisdictions for a case have agreed to share data, then race/ethnicity will be shown.

32 Recency data was determined using STARHS (Serologic Testing Algorithm for Recent HIV Seroconversion) results, when available, in combination with CD4-based staging data. If all the reporting jurisdictions for a case have agreed to share data, then recency data, will be shown.

33 Indicates if person was Stage 3 (AIDS) at HIV diagnosis.

34 Date of last negative HIV test is based on self-reported testing and treatment history (TTH) information or a lab documented negative HIV test before HIV diagnosis. If dates from both sources are available, the more recent date is chosen.

35 Most recent viral load (copies/mL) was based on national data. If all the reporting jurisdictions for a case have agreed to share data, then the most recent viral load result will be shown.

36 Date of the most recent viral load was based on national data. If all the reporting jurisdictions for a case have agreed to share data, then the corresponding date of the most recent viral load result will be shown.

37 Drug resistance was determined using Sierra, the Stanford HIV Web Service (Version 1.1).

38 Indicates the number of drug resistant-associated mutations were identified.

39 Indicates the specific mutations associated with nucleoside reverse transcriptase inhibitors (NRTIs), non-nucleoside reverse transcriptase inhibitors (NNRTIs), and protease inhibitors (PIs).

Appendix F. Suggested variables to be captured during a cluster investigation

As described in the roadmap to investigating and intervening in transmission clusters (see Figure 5), the first step of an investigation is to ascertain the known transmission cluster and risk network. To do so, first identify all named partners of molecular cluster members, capturing a few critical pieces of information as follows.

Table F1. Named partners of molecular cluster members.

Molecular cluster member			Named partner*			
Name	**StateNo**	**Partner services ID**	**Name**	**StateNo (if positive)**	**Partner services ID**	**HIV status (pos, neg, unk)**

*Consider also including persons who named the molecular cluster member, even if not named by the molecular cluster member.

This table can then be re-created to identify the partners of named HIV-positive partners (i.e., possible cluster cases and additional members of the risk network), as follows.

Table F2. Named partners of transmission cluster members.

Transmission cluster member (i.e., named HIV-positive partner)			Named partner*			
Name	**StateNo**	**Partner services ID**	**Name**	**StateNo (if positive)**	**Partner services ID**	**HIV status (pos, neg, unk)**

*Consider also including persons who named the transmission cluster member, even if not named by the transmission cluster member.

Once the known transmission cluster and risk network have been ascertained, the next step is to extract and review existing data for these persons. A list of suggested variables is below. Please note that, depending on the specific fields collected in a jurisdiction, some of these variables may be easily extracted from existing data sets, while others may only be captured through review of partner services comments fields and notes.

Table F3. Suggested variables for molecular cluster members, other transmission cluster members (i.e., HIV-positive partners), and other risk network members (i.e., HIV-negative or status unknown partners).

Variable	Molecular cluster members	Other transmission cluster members (i.e., HIV-positive partners)	Other risk network members (i.e., HIV-negative or status unknown partners)
StateNo	X*	X*	
Partner services ID	X*	X*	X*
Other reporting jurisdictions (per RIDR and CDC cluster snapshot)	X	X	X
Category (confirmed, probable, possible, risk network member)	X	X	X
Last name	X*	X*	X*
First name	X*	X*	X*
Aliases	X	X	X
DOB	X	X	X
Current address	X	X	X
Current city	X	X	X
Current county	X	X	X
Current state	X	X	X
Current zip code	X	X	X
Current address type	X	X	X
Census tract of residence	X	X	X
Country of birth	X	X	
Date of HIV diagnosis	X	X	
Facility at HIV diagnosis	X	X	
Residence at diagnosis	X	X	
Transmission category	X	X	
Sex at birth	X	X	X
Current gender	X	X	X
Age	X	X	X
Race/ethnicity	X	X	X
Acute HIV infection?	X*	X*	
Symptoms of acute HIV?	X*	X*	
If yes, date of symptom onset	X*	X*	
Recency (if available from STARHS testing)	X*	X*	
AIDS at diagnosis?	X	X	
Date of last negative HIV test (self report)	X*	X*	X*
Date of last documented negative HIV test	X*	X*	X*
Date of first CD4/viral load after diagnosis (Linkage)	X	X	
Most recent care facility	X*	X*	
Most recent VL value	X*	X*	
Most recent VL date	X*	X*	
Genotype in eHARS? (Y/N)	X	X	
Whether drug resistance detected	X	X	
Specific drug resistance mutations	X	X	

PS interview ever completed?	X*	X*	X*
Date of most recent PS interview	X*	X*	X*
Named partners (list StateNo or partner services ID	X	X	
Named social network members (list StateNo or partner services ID)	X	X	
Number of partners initiated	X*	X*	
Anonymous partners	X	X	
Number of partners located	X	X	
Number of partners tested	X	X	
Number of partners positive	X	X	
Number of social contacts named	X	X	
Number of partners last 12 months	X	X	X
Sex while drunk/high	X	X	X
Sex for drugs/$	X	X	X
Any recreational drugs	X	X	X
Any IDU	X	X	X
Meth use?	X	X	X
Sex with IDU			X
Sex with males			X
Sex with females			X
Sex with MSM			X
Sex with transgender persons			X
History of travel to other jurisdictions	X	X	X
Reported sex partners from other jurisdictions	X	X	X
Sex without condom	X	X	X
Reported places to meet partners	X	X	X
History of incarceration	X	X	X
Employment Status	X	X	X
History of military service	X	X	X
Current student	X	X	X
On PrEP prior to diagnosis	X	X	
# of STIs in the last 12 months	X	X	
Prior syphilis infection	X	X	X
Last syphilis date	X	X	X
Prior gonorrhea infection	X	X	X
Last gonorrhea date	X	X	X
Prior chlamydia infection	X	X	X
Last chlamydia date	X	X	X
Partner services disposition			X*
Partner services test result			X*
Referred for PrEP			X*
On PrEP			X*
Notes of pertinent other findings	X	X	X

*Indicates variables from the above data sets that can be extracted to prioritize and initiate persons for linkage to care and partner services.